RADIATION HYBRID MAPPING OF BARLEY CHROMOSOME 3H

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Mona Mazaheri

In Partial Fulfillment
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Plant Sciences
Option: Plant Breeding and Genetics

July 2014

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Radiation hybrid mapping of barley chromosome 3H

By

Mona Mazaheri

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Mohamed Mergoum
Co-Chair

Dr. Shahryar Kianian
Co-Chair

Dr. Michael Christoffers

Dr. Jane Schuh

Approved:

July 2, 2014
Date

Dr. Richard Horsley
Department Chair

ABSTRACT

Assembly of the barley (*Hordeum vulgare* L.) genome requires high resolution maps for aligning contig-based physical maps along chromosomes. Genetic maps lack accurate information on the physical position of almost half of the barley genome located in recombination-poor regions. Radiation hybrid (RH) mapping is an alternative approach, which is based on radiation-induced chromosomal deletions. In this study, an RH population for barley chromosome 3H was developed. Genotyping 373 3H-RH lines with 113 markers resulted in an RH map with an average resolution of 2.22 Kb. Compared to an analogous genetic map, the 3H-RH map resolution was 9.53-X higher, reaching to >262.40-X better resolution in the centromeric region. We suggest that RH maps would facilitate assembly of the barley genome. For future RH studies of the barley genome, an optimum genotyping platform, consisting of 400,536 barley-specific repeat junction markers (RJMs), was developed.

ACKNOWLEDGMENTS

My special thanks go to my advisors Dr. Shahryar Kianian, Dr. Mohamed Mergoum, and Dr. Penny Kianian for their support and valued advice through this research and also to my committee members Dr. Mike Christoffers and Dr. Jane Schuh. My thanks are extended to Allen Peckrul, Justin Hegstad, members of the former Wheat Germplasm Enhancement group, especially Dr. Monika Michalak de Jimenez and Dr. Ajay Kumar for their extensive help through this work. Also, I would like to thank my friends in the Plant Sciences Department and Reza who have been there for me and made the last four years an unforgettable memory.

I am also grateful for the "Frank Bain Dissertation Fellowship" and the "Charles and Linda Moses Presidential Graduate Fellowship" from College of Agriculture, Food Systems, and Natural Resources at North Dakota State University.

DEDICATION

To my parents and sister who were in my heart every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP…… Amplified fragment length polymorphism

BAC…… Bacterial artificial chromosome

bp……. Base pair

CS …… Chinese Spring

CS+3H" ……. Chinese Spring cultivar containing disomic addition barley chromosome 3H

cR…… Centi-Ray

DNA …… Deoxyribonucleic acid

EBC …… EST-based contigs

EST…….. Expressed sequence tag

Gb …… Giga base

GBS…… Genotyping by sequencing

Gc…… Gametocidal

Mb ……... Mega base

IBSC…… International Barley Genome Sequencing Consortium

IRAP ……... Inter-retrotransposon amplified polymorphism

ISBP …… Insertion-site-based polymorphism

Kb …… Kilo base

MAS …… Marker-assisted selection

Mid……... Middle

NHEJ ……... Non-homologues end joining

RBIP …… Retrotransposon-based insertion polymorphism

RFLP…… Restriction fragment length polymorphism

RH …….. Radiation hybrid

RJ……. Repeat junction

RJJM ……... Repeat junction-junction marker

RJM ……. Repeat junction marker

RR ……... Random reads

SNP…….. Single nucleotide polymorphism

SSR……..... Single sequence repeat

STS……... Sequence tagged sites

TE …… Transposable element

TREP ……... Triticeae repeat

YAC…… Yeast artificial chromosomes

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

1. INTRODUCTION AND LITERATURE REVIEW

1.1. Importance of Barley

Barley (*Hordeum vulgare* L.), with a global production of 133 million tons per year, ranks fourth among cereal crops, after maize (*Zea mays* L.), rice (*Oryza sativa* L.), and bread wheat (*Triticum aestivum* L.) (FAO, 2012, http://faostat.fao.org). In the United States, barley has an annual production of ~4.7 million tons, cultivated on 1.3 million hectares (FAO, 2012, http://faostat.fao.org). About 75% of the barley production is used for animal feed, 20% in beer brewing, and only 5% in human diets (Blake et al., 2011). Barley, however, is the main food source in poor countries because of the adaptation to harsh environmental conditions such as cold, drought, alkali, and saline soils (Grando and Macphereson, 2005; Schulte et al., 2009). In developed countries, barley is considered as a super food due to substantial health benefits such as high β-glucan content (2.0-20.0 g/100 g flour in barley compared to 0.5-6.2 g/100 g flour in other cereals) (Rudi et al., 2006, Collins et al., 2010). β-glucan reduces the risk of diet-related diseases such as type II diabetes and colorectal cancer (Collins et al., 2010).

Some researchers use barley as a model plant in Triticeae genomic studies. The Triticeae tribe evolved ~12 million years ago within the Pooideae subfamily of Poaceae and includes barley, wheat (*T. aestivum* L., *T. durum* L.), rye (*Secale cereale* L.), and their wild relatives (Gaut, 2002). The Triticeae genome is large with high content of repetitive sequences (~80% of the nuclear DNA) (Dvorak, 2009). A model plant would facilitate studying the complex Triticeae genome. Among the most popular grass models (rice, maize, and Brachypodium [*Brachypodium distachyon* L.]), Brachypodium has the closest phylogenetic relationship to the Triticeae species (Vogel and Bragg, 2009). This relationship; however, is not close enough to extrapolate the majority of Triticeae genetic information from the Brachypodium genome. For example,

common wheat and Brachypodium diverged about 35-40 million years ago (Bossolini et al., 2007), and only less than one third of the wheat genes are in common with Brachypodium (Kumar et al., 2009). On the contrary, gene sequences are highly conserved within Triticeae species (Van Deynze et al., 1995; Dubkovsky et al., 1996). For example, genic sequences of barley had 93% similarity with that of wheat (Sato et al., 2009). A model plant within this tribe would best represent the Triticeae genome. Due to its diploid genome (2n=14) and its self-pollinated characteristic, barley is considered as a model plant in genomic studies of Triticeae crops (Schulte et al., 2009). In barley, identifying genes underlying the vernalization requirement (*VRV-1*) (Fu et al., 2005) and the photoperiod response (*Ppd-H1*) (Turner et al., 2005) led to the detection of corresponding genes in common wheat (Beales et al., 2007; Cockram et al., 2007).

1.2. Barley Domestication

Barley was domesticated about 10,000 years ago in the Fertile Crescent from the wild relative, *H. spontaneum* C Koch (Bard et al., 2000). Comparing amplified fragment length polymorphism (AFLP) patterns of wild barley varieties, collected around the world, with that of domesticated cultivars revealed the Israel-Jordan area as the barley domestication site. During the migration of domesticated barley from the Fertile Crescent to South Asia, several allelic substitutions on a locus (*BKn-3* gene) occurred in the Himalayas. Therefore, the Himalayas is known as the diversification site of domesticated barley (Kilian et al., 2009). Compared to the wild type, cultivated barley has shorter stems, spikes, and awns, broader leaves, and larger grains (Zohary 1969). Barley varieties are two- or six-rowed, based on morphology of the spikelet. Wild barley is two-rowed, whereas cultivated barley is either two- or six-rowed, with six-rowed cultivars having wider distribution than two-rowed. The phylogeny of two and six-rowed varieties is controversial. Some studies suggested that barley domestication was diphilic; that is,

two and six-rowed had two different origins (Komatsuda et al., 2007; Kilian et al, 2006). In contrast, some studies proposed a same origin for both variants; where six-rowed genotypes derived from two-rowed by mutations at the *vrs1* locus (Badr et al., 2000; Tanno et al., 2002).

1.3. Current Status of Barley Genetic and Genomic Resources

The barley genome is 5.1 Gb in size and consists of ~84% repetitive elements (Mayer et al., 2012). Complexity of the barley genome makes genomic studies challenging. During the last two decades a large number of barley genomic resources have been developed to facilitate analyzing the structure and function of the genome. These resources include large collections of natural and mutant lines, comprehensive expressed sequence tag (EST) data, and numerous genetic and physical maps.

1.3.1. Collections of natural and induced genetic diversity

Large numbers of barley wild relatives, landraces, traditional and modern cultivars are collected around the world and are stored in several genebanks. These resources are extensively used by researchers to discover new traits and genes underlying those traits. In 2009, genebank collections were comprised of 453,602 genotypes in which the majority belonged to cultivated barley (*H. vulagre* subsp. *vulgare*) accessions and its close wild relatives (*H. vulagre* subsp. *spontaneum*) (Knupffer, 2009). The largest barley genebank centers are located in Canada (8.8% of the total stored accessions), USDA (6.7%), Brazil (6.4%), and ICARDA (5.7%). (Knupffer, 2009). Testing a subset of 95,000 barley accessions from major genebank centers showed that 85% of the global stored accessions were unique and stored in only one of the genebank centers (Hintum and Menting, 2003).

To extend barley genetic diversity beyond the natural variation, thousands of radiation-mediated mutants were developed and characterized during the last 80 years (summarized in

http://wheat.pw.usda.gov/ggpages/bgn/26/bgn26tc.html). Barley mutants are utilized to discover various genes underlying physiological (Morell et al., 2003; Clarke et al., 2008) and morphological traits (Komatsuda et al., 2007, Taketa et al., 2008, Chono et al, 2003)

1.3.2. EST data

A comprehensive atlas of the barley transciptome, comprising of about half-million ESTs, has been developed (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html, January 1st 2013 release). These ESTs were generated from different tissues of several genotypes and cover genes underlying various physiological traits such as plant development and response to biotic and abiotic stresses. Analyzing EST data revealed approximately 26, 593 non-redundant genes per barley genome (Sreenivasulu et al., 2008). Among barley unigenes, 92.3% were genetically mapped with the gene density averaging five genes per Mb (Mayer et al., 2012). Generated ESTs were utilized to develop several microarray platforms for functional genomic studies. One example of barley gene-based arrays is the Affymetrix microarray containing 21,349 unigenes (Close et al., 2004), which was applied to detect gene networks underlying tolerance to drought (Ozturk et al., 2002) and cold stresses (Faccioli et al., 2002). EST data were also applied for designing PCR-based markers and generating highly saturated genetic maps. Of particular note is the generation of 1,000-EST (Stein et al., 2007) and a 30,000-EST barley genetic maps (Sato et al., 2009), which are valuable sources for detecting candidate genes underlying agronomically important traits.

1.3.3. Genetic maps

Genetic maps estimate the location of loci on a chromosome based on the recombination rate; loci with low recombination rates are mapped at closer distances, compared to loci with high recombination rates. Since the first barley genetic map, consisting of 115 restriction

fragment length polymorphism (RFLP) markers (Heun et al., 1991), numerous genetic maps with various molecular markers have been developed (http://wheat.pw.usda.gov/ggpages/map_summary.html) such as a genetic map with 1,172 AFLP, single sequence repeat (SSR), and sequence tagged sites (STS) (Hori et al., 2003). Mapping information was further enhanced by integrating several genetic maps and developing consensus maps for the barley genome. For instance, by combining datasets of ten recombinant populations, a consensus genetic map with ~3000 diversity arrays technology (DArT), SSR, RFLP, and STS markers was developed (Wenzl et al., 2006).

The advent of next generation sequencing techniques enabled simultaneous genotyping of several lines with thousands of markers and accelerated developing high-density genetic maps. As a case in point, the single nucleotide polymorphism (SNP)-based Illumina GoldenGate BeadArray genotyping platform resulted in a barley genetic map with 2,943 SNPs (Close et al., 2009). Using genotyping by sequencing (GBS) technique, a genetic map consisting of 34,000 SNPs and 240,000 tags (Poland et al., 2012) was constructed.

The drawback of recombination-based genetic maps is that they do not reveal the exact physical location of mapped loci. In Triticeae, the recombination rate is almost zero in the centromeric region and increases toward telomeres by nearly square of the relative distance from the centromere (Kunzel et al., 2000; Akhunov et al., 2003). Due to the uneven recombination distribution, large variation in genetic to physical distance is reported in the barley genome, ranging from 1.5 Mb/cM in distal to 89 Mb/cM in proximal regions (Stephens et al., 2004). Moreover, the recombination rate is higher among closely related genotypes, compared to distant genotypes (Peters et al., 2003). Therefore, the genetic distance between some loci might change by using different recombinant populations. Another disadvantage of genetic maps is that about

half of the barley genome, located in recombination-poor regions, is inaccessible on genetic maps, whereas recombination-rich regions account for only 4.9% of the barley genome (Kunzel et al., 2000).

1.3.4. Physical maps

Physical maps determine the physical locations of loci on chromosomes. Different types of physical maps are available, which vary in the map resolution. Using addition lines of bread wheat cultivar "Chinese Spring" (CS), disomic for barley (*H. vulgare* L.) cultivar "Betzes" chromosomes, 1,787 ESTs were assigned to barley chromosomes (Cho et al., 2006). For assigning markers to chromosomal regions, gametocidal (Gc) maps, also known as deletion bin maps, were developed. The addition of an Aegilops (*Aegilops cylindrical* L.*)* chromosome, known as the Gc chromosome, in a bread wheat induces terminal deletions in the bread wheat chromosomes and also in chromosomes added to the wheat background such as barley chromosomes (Endo et al., 1988; Endo 2007). By integrating the Gc chromosome into wheat-barley addition lines, cytological stocks of the barley genome were developed (Shi and Endo, 1997, 1999). Through these stocks, several Gc maps were developed for the barley chromosome 2H (Giri et al., 2011), 3H (Sakai et al., 2009), 4H (Sakata et al., 2010), 5H (Ashida et al., 2007), and 7H (Serizawa et al., 2001). The drawback of Gc maps is that they do not separate loci located at proximal regions. Moreover, Gc maps do not order markers within single deletion bins.

Contig-based physical maps have the highest map resolution and reflect the exact physical distance between loci; however, generating these maps is a tedious process. First, genomic DNA is digested into fragments with particular lengths. DNA fragments are then cloned by inserting into a vector (bacterial artificial chromosome [BAC] or yeast artificial chromosome

[YAC]) leading to the development of genomic libraries. After fingerprinting the inserted fragments, overlapped clones are assembled into contigs. As the last step, contigs are anchored onto chromosomes using genetic maps. Contig-based physical maps are used for the map-based cloning of important genes and also as a template for clone by clone sequencing of the genome. A genome-wide physical map of barley was developed through generation of 571,000 BACs from six different libraries (Schulte et al., 2011). Assembly of the BACs led to 9,265 contigs, covering ~98% of the barley genome (Mayer et al., 2012).

1.4. Barley Genome Sequencing

Despite advances in generating genomic resources, to date only few barley genes have been cloned (for a review see Krattinger et al., 2009). A complete barley genome sequence enables pinpointing the exact physical locations of genes and would ease the long process of map-based cloning. Moreover, a high quality reference genome reveals gene networks and regulatory mechanisms underlying important traits. This information would be valuable in breeding programs of barley and related Triticeae crops. To sequence the barley genome, the International Barley Genome Sequencing Consortium (IBSC; http://barleygenome.org) was established in 2006 with the collaboration of six countries (For a review see Schulte et al., 2009). Despite international efforts, high complexity of the barley genome precludes development of a complete genome sequences. Based on a recent IBSC update, *de novo* assembly of barley deep shotgun sequencing (50X coverage) resulted in assembly of contigs totaling 1.9 Gb in length. Remainder of the genome (3.2 Gb) was not assembled due to the high portion of repetitive elements (Mayer et. al., 2012). An alternative approach for assembly of complex genomes is aligning contig-based physical maps by anchoring them onto genetic maps. Taking this approach, 1.2 Gb of the barley genome remained unassembled (Mayer et al., 2012), despite

applying a high saturated genetic map with approximately half a million markers (Polland et al., 2012). This result was expected because contigs located in recombination-poor regions cannot be anchored onto genetic maps due to the low map resolution of these regions.

1.5. Radiation Hybrid Maps

An alternative to conventional genetic mapping is radiation hybrid (RH) mapping. In this method, random deletions are induced along all chromosomal regions by applying ionizing radiation, such as X and gamma-rays. Chromosomes with induced deletions are rescued by fusion into non-native recipient cells. Frequency of the breakage between loci is used to estimate the relative position of each locus to other loci. Markers with small physical distance will show less breakage, compared to physically distant markers. The RH map unit is the centi-Ray (cR), which corresponds to one break between two loci in every 100 lines. The main advantage of RH maps is that they cover all chromosomal regions. Also, the map resolution can be increased by inducing more deletions via applying higher radiation dosage. Moreover, since markers are mapped based on presence vs. absence, both polymorphic and monomorphic markers can be used for genotyping an RH population.

1.5.1. Physiological mechanisms underlying radiation-induced deletions

A general model is proposed to explain mechanisms underlying inducement of radiation-mediated deletions. During exposure of a biologic system to ionizing radiations, cells randomly receive the radiation energy (Kumar et al., 2014). Since water composes a majority of the cell volume (80%), the major effect of radiation is evident on the cell water (Britt 1996). Radiation energy decomposes the cell water into reactive oxygen species (ROS), such as superoxide anion (O_2^-) and hydroxyl radicals (OH) (Wi et al., 2007). ROC oxidates the sugar phosphate group of DNA molecules, which results in breaking phosphodiester bonds and inducing double strand

breaks (DSB) within chromosomes. Exonuclease enzymes then degrade nucleotides at the breakage point and expand the induced deletion (Britt 1996). In somatic cells the broken strands are rejoined by non-homologues end joining (NHEJ) DNA repair mechanism. In this mechanism a protein complex (*Ku* protein) attaches to the breakage site to prevent further DNA degradation and to pull the broken strands together. The broken ends are then rejoined by DNA ligase activity. During the repair mechanism, degraded nucleotides at the original DSB site are not restored, which leads to deleted segments within chromosomes (Bleuyard et al., 2006).

Applying a high radiation dosage maximizes the number and size of induced deletions (Hlatky et al., 2002). In cotton (*Gossypium hirsutum* L.), a 5 Krad gamma-ray induced greater numbers of deletions compared to a 1.5 Krad gamma-ray (Gao et al., 2004, 2006). The deletion frequency of the common wheat D genome increased by 1.2% by increasing gamma-radiation dosage from 15 to 45 Krad (Kumar et al., 2012). In wheat-barley chromosome 7HS addition lines, increasing the gamma-radiation dosage from 15 to 25 Krad inhibited plant growth and reduced the spike length (Figure 1.1). The reduced vigor of wheat-barley chromosome 7HS addition lines treated with a high radiation dose could be related to the large deletion frequency (Kumar et al., 2014).

In generating an RH population, applying different radiation dosages induces deletions with various sizes within the population. Large induced deletions can overlap toward generating a minimal tilling path to cover a whole chromosome, whereas small deletions lead to separating physically close loci. Thus, RH lines containing various sized-deletions enable generating contiguous RH maps with high resolution (Stewart et al. 1997). The tissue moisture content also affects the deletion inducement. The deletion frequency increases with moisture due to the inducement of high levels of ROS that invade chromosomes (Kumar et al., 2014).

Figure 1.1. Effect of 15 and 25 Krad gamma-radiations on the spike length of wheat-barley chromosome 7HS addition lines. Spike 1, 2, and 3 were sampled from lines irradiated with 25Krad and spike 4 and 5 from lines irradiated with 15 Krad. Spike 6, 7 and 8 were control samples collected from lines with no radiation treatment.

1.5.2. History of RH mapping

The RH mapping technique was first described by Goss and Harris in 1975 by irradiating human X chromosomes and fusing cells containing fragmented chromosomes with non-irradiated recipient rodent cells. This method was later improved to develop an RH population of the whole human genome (Walter et al., 1994). Through this panel, a human wide-genome RH map with ~41,000 STS was constructed. The generated map had 100 kb resolution and covered 30,000 human unigenes (Stewart et al.,1997). Following the success of the RH method in mapping the human genome, this method was adapted to map large numbers of mammalian genomes (For review, see Faraut et al., 2009).

Compared to animals, polyploid plants are more tolerant to genomic changes such as introgression of alien chromosomes (Kumar et al., 2014). The genome plasticity allows for

developing viable addition and substitution lines for polyploids. These cytological stocks can replace the process of *in vitro* cell fusion and lead to *in vivo* RH populations. Through viable RH panels, induced deletions can be retained in the next generation for further analysis. Also, *in vivo* RH populations can be utilized in forward genetic studies to associate mutant phenotypes with deleted DNA fragments (Michalak de Jimenez et al., 2013). The first plant RH population was developed for maize (*Zea mays* L.) chromosome 9 (Riera-Lizarazu et al., 2000) by crossing oat (*Avena sativa* L.) lines carrying maize disomic addition chromosome 9, with normal oat. Resulting seeds, which were monosomic for maize chromosome 9, were irradiated with 30, 40, and 50 Krad gamma-rays. Irradiated seeds were planted and mature plants were self-pollinated. Lines resulting from self-pollination of irradiated seeds formed the maize chromosome 9-RH population, which contained homozygous deletions on this chromosome. Screening the population revealed 1-10 induced deletions per line randomly distributed along the chromosome. It was estimated that 100 RH lines with an average of three breakages per chromosome would result in an RH map with 0.5 to 1 Mbp resolution. Using the same method, 171 RH lines of maize chromosome 1 were generated for constructing an RH map with 45 markers (Kynast et al., 2004).

In common wheat, an RH population of wheat chromosome 1D was developed using substitution durum lines, in which chromosome 1A was replaced with chromosome 1D (27"+1D") (Hossain et al., 2004). The seeds of substitution lines were irradiated with 35 Krad of gamma-rays, and matured lines were crossed to normal durum lines (AABB, 2n=2x=28) to generate wheat chromosome 1D-RH population containing the addition of a single 1D chromosome with induced deletions. Using this population an RH map with 39 markers was constructed. The same panel was used to develop an RH map with a higher density, consisting of

378 molecular markers (Kalavacharla et al., 2006). Resolution of the generated map was estimated to be ~199 Kb. This level of resolution allowed for anchoring several BAC contigs onto the RH map. Among plant RH maps, wheat chromosome 3B-RH map, with 541 markers and 90 Kb resolution (Kumar et al., 2012), had the highest map resolution. For generating 3B-RH populations, seeds of normal durum were irradiated with 35Krad gamma-radiation. Matured irradiated lines were then crossed with substitution durum lines with chromosome 3D in place of chromosome 3B (13"+3D").

Plant single chromosome-RH populations were extended to RH populations for a whole genome. A cotton whole-genome RH panel was developed by fertilizing eggs of a cultivated cotton (*Gossypium. Barbadense* L.*)* with irradiated pollens of a related species (*G. hirsulum* L.) (Gao et al., 2004). Through this panel, 102 markers were mapped, and several markers previously assigned to genetic maps were reassigned to new genomic locations. Recently 1,500 RH lines of common wheat D genome were developed by crossing irradiated hexaploid wheat (AABBDD), treated with 35 and 45 Krad, to durum wheat (AABB) (Kumar et al., 2012). Mapping efficiency of generated populations was tested by mapping 15 markers on the wheat chromosome 2D.

In barley, an *in vitro* whole genome RH population was developed by fusing barley protoplasts, irradiated with 5 Krad of X-ray, with untreated tobacco (*Nicotiana tabacum*) protoplasts (Wardrop et al., 2002). The barley protoplasts contained a transgene conferring resistance to bialaphos. RH hybrids were selected by culturing them in a bialaphos-containing media. Using this technique 40 barley RH calli were developed, and the inducement of deletions was confirmed by genotyping the calli with 35 markers. Due to the lack of enough tissue, RH lines were not further genotyped.

1.5.3. RH maps in genomic studies

RH maps cover all chromosomal regions and have a higher resolution, compared to genetic maps. Therefore, saturated RH maps were extensively used as a platform for assembly of human and animal genomes (Lander et al. 2001; Waterston et al., 2002; Gibbs et al., 2004; Lindblad-Toh et al. 2005; Wade et al. 2009). To date the majority of plant RH maps are in preliminary stages; however, they have confirmed inducement of deletions along all chromosomal regions. The efficiency of RH maps in plant genome assembly was tested by applying the wheat chromosome 3B-RH map for contig assembly of the corresponding chromosome sequences (Paux et al., 2008).

Although RH maps facilitate aligning contigs, the development of contigs is a heavy workload that requires costly genome deep sequencing and laborious fingerprinting of thousands of clones. An alternative to the whole genome sequencing is development of comprehensive saturated RH maps. In this approach, short sequence reads of survey sequencing data (1-2X coverage) are anchored onto RH maps for further analysis such as functional genomics, comparative genetics, and evolutionary studies (Hitte et al., 2005). Taking this approach, although the generated genome sequence is not contiguous, it provides a general view of the genome structure at a low cost and within a short time frame. Once an RH population is developed, numerous markers can be mapped through PCR amplification or high throughput genotyping platforms for saturating the genome. In canine, a dense RH map with 200 Kb-300 Kb resolution provided the same genomic information for ~10% of the cost of whole-genome sequencing (6-8X coverage) (Hitte et al., 2005).

1.6. Repeat Junction Markers in RH Mapping

For genotyping an RH population, markers need to be genome-specific so that they do not interfere with the recipient genome. Considering the conserved synteny among species genomes, especially among genic sequences of diverse species (Timms et al., 2006), generating markers that are polymorphic between donor and recipient genomes is challenging, particularly in the case of designing gene-based markers. For example, among ~10,000 EST markers designed from the barley genome, 8,579 (85.79%) were in common with the common wheat genome (Nasuda et al., 2005). One solution for designing genome-specific markers would be aligning the donor and host genomes and designing markers based on sequences that differ between the two genomes. Designing markers with this approach, however; would be time-consuming and requires availability of the sequences of both donor and host genomes.

Another solution for designing genome-specific markers is applying repeat junction (RJ) sequences of transposable elements (TEs). TEs are expanses of repeat sequences that translocate within a genome. Based on the mode of translocation, TEs are divided into two major classes: class I, or retrotransposons, and class II, or DNA transposons. Retrotransposons translocate by amplifying their sequences and inserting the copied element into a different region. In contrast, DNA transposons change their genomic location by excising from one region and inserting into a new location. In both classes, TE insertions generate unique RJ oligonucleotides consisting of boundaries between TEs and sequences of which TEs are inserted. RJ sequences are unique, in that; they are genome-specific, and each RJ exists as a single copy within a genome (Bennetzen et al., 2000; You et al., 2010). Therefore, markers designed from RJ sequences would be unique and genome-specific. Moreover, since TEs constitute >80% of the Triticeae genome, distributed

evenly along chromosomes (Sabot and Schulman, 2009), RJ based markers could be abundant with a high genome-coverage.

To date several RJ-based markers have been developed such as inter-retrotransposon amplified polymorphism (IRAP), retrotransposon-based insertion polymorphism (RBIP), insertion-site-based polymorphism (ISBP), repeat junction markers (RJM), and repeat junction-junction markers (RJJM). RJ-based markers can be obtained from whole genome sequences through an automated high throughput design by "RJPrimer" software (You et al., 2010). Using this software, from the rice chromosome 1, a total of 4,235 RJ-markers were designed consisting of 1,897 RJMs, 28 RJJMs, 1,837 ISBPs, 443 RBIPs, and 30 IRAPs (You et al., 2010).

Designing IRAP, RBIP, and ISBP markers is similar; in that, one primer (forward or reverse) is designed from the margin sequences of a TE. The second primer is designed from either the margin sequences of a different TE (in the IRAP system) (Kalendar et al., 2006), or from a genic sequences (in the RBIP system) (Flavell et al., 1998), or from any genomic sequence such as TEs or genes (ISBP markers) (Paux et al., 2006). The three above-mentioned markers produce amplicons containing genome-specific RJ sequences, but these markers are not specifically designed from the RJ sequences. Therefore, it is likely that these three markers produce same-sized amplicons from different genomes. Thus, IRAP, RBIP, and ISBP markers do not necessarily differentiate between genomes by the simple technique of PCR amplification and electrophoresis gels.

To differentiate genomes based on presence/absence of single amplicons, RJMs and RJJMs were developed. In both marker systems, one primer is designed from an RJ sequence. In RJJMs the other primer spans a different RJ sequence (Luce et al., 2006), whereas in RJMs the other primer could be picked from any genomic sequence (Wanjugi et al., 2009). Due to the high

flexibility of RJM design, the frequency of RJMs within a genome is considerably higher than that of RJJMs. For example, from rice chromosome 1 sequences, the quantity of designed RJMs (1,897) was ~68-fold higher than the quantity of designed RJJMs (28) (You et al., 2010).

The genome specificity and abundance of RJMs make them an ideal platform for genotyping RH populations. The genome specificity of RJMs was confirmed by testing RJMs of wheat D-genome progenitor, *Ae. tauschii* (DD), against the three genomes in common wheat (AABBDD). Among 269 RJMs designed from *Ae. tauschii*, 97% specifically amplified the wheat D genome without amplifying the A and B genomes (Wanjugi et al., 2009). RJMs are also reported to be evenly distributed along genomes, which is ideal for detecting radiation-mediated deletions induced throughout chromosomes (You et al., 2010; Kumar et al., 2013). In screening a wheat D genome-RH population, RJMs detected 8.9% of induced deletions, whereas SSRs and ESTs detected 3.8% and 3.2% of induced deletion, respectively (Kumar et al., 2013).

1.7. Objectives

Objectives of this study were I) generating an *in vivo* barley RH population; II) constructing an RH map for barley; and III) optimizing a genotyping platform for studying the RH population. The next chapter of this thesis explains the development of an RH panel and an RH map for a barley chromosome (chromosome 3H). Advantages of the generated RH map are also elucidated by comparing it to an analogous genetic map. The third chapter describes the development of large numbers of barley-specific RJMs and the advantages of this marker in RH mapping and analyzing the barley genome.

1.8. References

Akhunov, E.D., A.W. Goodyear, S. Geng, L.L. Qi, and B. Echalier *et al.* 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. 13:753-763.

Ashida, T., S. Nasuda, K. Sato, and T.R. Endo. 2007. Dissection of barley chromosome 5H in common wheat. Genes. Genet. Syst. 82:123-133.

Badr, A., K. Muller, R. Schafer, H. el Rabey, S. Effgen, H.H. Ibrahim, C. Pozzi, W. Rohde, and F. Salamini. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). Mol. Biol. Evol. 17(4):499-510.

Beales, J., A. Turner, S. Griffiths, J. Snape, and D. Laurie. 2007. A pseudoresponse regulator is misexpressed in the photoperiod insensitive *PpdD1a* mutant of wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 115:721-733.

Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. 42:251-269.

Blake, T., V. Blake, J. Bowman, and H. Abdel-Haleem. 2011. In: S.E. Ullrich, editor, Barley: Production, improvement and uses. Wiley-Blackwell. p.522-531.

Bleuyard, J.Y., M.E. Gallego, and C. I. White. 2006. Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA repair. 5:1-12.

Bossolini, E., T. Wicker, and P.A. Knobel. 2007. Comparison of orthologous loci from small grass genomes Brachypodium and rice: implications for wheat genomics and grass genome annotation. Plant J. 49:704-717.

Britt, A.B. 1996. DNA damage and repair in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:75-100.

Cho, S., D.F. Garvin, and G.J. Muehlbauer. 2006. Transcriptome analysis and physical mapping of barley genes in wheat-barley chromosome addition lines. Genetics. 172(2):1277-1285.

Chono, M., I. Honda, H. Zeniya, K. Yoneyama, D. Saisho, K. Taketa, S. Takatsuto, T. Hoshino, and Y. Watanabe. 2003. A semidwarf phenotype of barley results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol. 133:1209-1219.

Clarke, B., R. Liang, M.K. Morell, A.R. Bird, C.L. Jenkins, and Z. Li. 2008. Gene expression in a starch synthase *IIa* mutant of barley: changes in the level of gene transcription and grain composition. Funct. Integr. Genomics. 8:211-221.

Close, T.J., S.I. Wanamaker, R.A. Caldo, S.M. Turner, D.A. Ashlock, J.A. Dickerson, R.A. Wing, G.J. Muehlbauer, A. Kleinhofs, and R.P. Wise. 2004. A new resource for cereal genomics: 22K barley genechip comes of age. Plant Physiol. 134(3):960-968.

Close, T.J., P.R. Bhat, S. Lonardi, Y. Wu, N. Rostoks, and K. Ramsay *et al*. 2009. Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics. 10:582-595.

Cockram, J., H. Jones, F.J. Leigh, D. Osullivan, W. Powell, D.A. Laurie, and A.J. Greenland. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J. Exp. Bot. 58:1231-1244.

Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luo, M.J. Kearsey, P. Warner, D .Harrap, C. Tapsell, H. Liu, P.E. Hadley , N. Stein, D. Schulte, B. Steuernagel, D.F. Marshall, W.T. Thomas, L. Ramsay, L. Mackay, D.J. Balding, R. Waugh, and D.M. O'Sullivan. 2010. Genome-wide association mapping to candidate

polymorphism resolution in the unsequenced barley genome. Proc. Natl. Acad. Sci. USA. 107:21611-21626.

Collins, H.M., R.A. Burton, D.L. Topping, M.L. Liao, A. Bacic, and G.B. Fincher. 2010. Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: Potential importance in human health and nutrition. Cereal Chem. 87(4):272-282.

Druka, A., G. Muehlbauer, I. Druka, R. Caldo, U. Baumann, N. Rostoks, A. Schreiber, R. Wise, T. close, A. Kleinhofs, A. Graner, A. Schulman, P. Langridge, K. Sato, P. Hayes, J. McNicol, D. Marshall, and R. Waugh. 2006. An atlas of gene expression from seed to seed through barley development. Func. and Integ. Genomics. 6(3):202-211.

Dubcovsky, J., M.C. Luo, G.Y. Zhong, R. Bransteitter, A. Desai, A. Kilian, A. Kleinhofs, and J. Dvorak. 1996. Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983-999.

Dvorak, J. 2009. Triticeae genome structure and evolution. In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p. 451-477.

Endo, T.R. 2007. The gametocidal chromosome as a tool for chromosome manipulation in wheat. Chromosome Res. 15:67-75.

Endo, T.R. and B.S. Gill. 1996. The deletion stocks of common wheat. J. Heredity. 87:295-307.

Erayman, M., D. Sanduh, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res. 32:3546-3565.

Faccioli, P., M.S. Lagonigro, L. De Cecco, A.M. Stanca, R. Alberici, and V. Terzi. 2002. Analysis of differential expression of barley ESTs during cold acclimatization using microarray technology. Plant Biol. 4:630-639.

Faraut, T., S. de Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, T. Schiex, B. Servin, A. Vignal, F. Galibert, and M. Yerle. 2009. Contribution of radiation hybrids to genome mapping in domestic animals. Cytogen. And Genome Res. 126:21-33.

Faris, J.D., J.P. Fellers, S.A. Brooks, and B.S. Gill. 2003. A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics. 164:311-321.

Feuillet, C. and J. Salse. 2009. Comparative genomics. In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p. 451-477.

Flavell, A.J., M.R. Knox, S.R. Pearce, and T.H. Ellis. 1998. Retrotransposon-based insertion polymorphism (RBIP) for high throughput marker analysis. Plant J. 16:643-650.

Fu, D., P. Szucs, L. Yan, M. Helguera, J.S. Skinner, J. van Zitzewitz, P.M. Hayes, and J. Dubcovsky. 2005. Large deletions within the first intron in *VRV*-1are associated with spring growth habit in barley and wheat. Mol. Genet. Genomics. 273:54-65.

Gao, W., Z.J. Chen, J.Z. Yu, R.J. Kohel, J.E. Womack, and D.M. Stelly. 2006. Wide-cross whole-genome radiation hybrid mapping of the cotton (*Gossypium barbadense* L.) genome. Mol. Gen. Genomics. 275:105-113.

Gao, W., Z.J. Chen, J.Z. Yu, D. Raska, R.J. Kohel, J.E. Womack, and D.M. Stelly. 2004. Wide-Cross Whole-Genome Radiation Hybrid Mapping of Cotton (*Gossypium hirsutum* L.). Genetics. 167:1317-1329.

Gaut, B.S. 2002. Evolutionary dynamics of grass genomes. New Phytol. 154:15-28.

Gibbs, R.A., G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, and S. Scherer *et al*. 2004. Genome sequence of the brown Norway rat yields insights into mammalian evolution. Nature. 428:493-521.

Joshi, G.P., S. Nasuda, and T.R. Endo. 2011. Dissection and cytological mapping of barley chromosome 2H in the genetic background of common wheat. Gene Genet. Syst. 86:231-248.

Goss, S.J. and H. Harris. 1975. New method for mapping genes in human chromosomes. Nature. 255:680-684.

Grando, S. and H.G. Macpherson. 2005. Food barley: Importance, uses, and local knowledge. ICARDA, Aleppo, Syria.

Hearnden, P.R., P.J. Eckermann, G.L. McMichael, M.J. Hayden, J.K. Eglington, and K.J. Chalmers. 2007. A gentic map of 1,000 SSR and DArT markers in a wide barley cross. Theor. Appl. Genetics. 115(3):383-391.

Heun, M., A.E. Kennedy, J.A. Anderson, N.L.V. Lapitan, M.E. Sorrells, and S.D. Tanksley. 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) Genome. 34(3):437-447.

Hintum, T.J.L.V. and F. Menting. 2003. Diversity in *ex situ* genebank collections of barley. In: R. Von Bothmer, T.J.L.V. Hitnam, H. Knupffer, and K. Sato, editors, Diversity in barley (*Hordeum vulgare*). Elsivier Science B.V. Amsterdam . The Netherlands. p.247-257.

Hitte, C., J. Madeoy, E.F. Kirkness, C. Priat, T.D. Lorentzen, and F. Senger. 2005. Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. Nature Rev. Genet. 6(8):643-648.

Hlatky, L., R.K. Sachs, M. Vazquez, and M.N. Cornforth. 2002. Radiation-induced chromosome aberrations: insight gained from biophysical modeling. Bioessays. 24:714-723.

Hori, K., T. Kobayashi, A. Shimizu, K. Sato, K. Takeda, and S. Kawasaki. 2003. Efficient construction of high-density linkage map and its application to QTL analysis in barley. Theor. Appl. Genet. 107:806-813.

Hossain, K.G., O. Riera-lizarazu, V. Kalavacharla, M.I. Vales, J.L. Rust , S.S. Maan, and S.F. Kianian. 2004a. Molecular cytogenetic characterization of an alloplasmic durum wheat line with a portion of chromosome 1D of *Triticum aestivum* carrying the scs^{ae} gene. Genome. 47:206-214.

Hossain, K.G., O. Riera-lizarazu, V. Kalavacharla, M.I. Vales, S.S. Maan, and S.F. Kianian. 2004b. Radiation hybrid mapping of the species cytoplasm-specific (scs^{ae}) gene in wheat. Genetics. 168:415-423.

Jannink, J.L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics. 9:166-177.

Kalendar, R. and A.H. Schulman. 2006. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat. Protoc. 1:2478-2484.

Kilian, B., H. Ozkan, J. Kohl, A. Haeseler Von, F. Barale, O. Deusch, A. Brandolini, C. Yucel, W. Martin, and F. salamini. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol. Gene. Genome. 276:230-241.

Knupffer, H. 2009. Triticeae genetic resources in *ex situ* genebank collections. In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p 31-79.

Komatsuda, T., M. Pourkheirandish, C. He, P. Azhaguvel, P. Kanamori, H. Perovic, N. Stein, A. Graner, T. Wicker, and A. Tagiri. 2007. Six –rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc. Natl. Acad. Aci USA. 104:1424-1429.

Krattinger, S., T. Wicker, and B. Keller. 2009. Map-based cloning of genes in Triticeae (Wheat and balrley). In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p 337-357.

Kumar, S., A. Mohan, H.S. Balyan, and P.K. Gupta. 2009. Orthology between genomes of *Brachypodium*, wheat, and rice. BMC Genomics. 2:93-102.

Kumar, A., F.M. Bassi, M. Michalak de Jimenez, F. Ghavami, M. Mazaheri, K. Simons, M.J. Iqbal, M. Mergoum, S.F. Kianian, and P.M.A. Kianian. 2014. Radiation hybrids: A valuable tool for genetic, genomic, and functional analysis of plant genomes. In: R. Tuberosa, A. Graner, and E. Frison, editors, Genomics of Plant Genetic Resources. Springer. Netherlands. p. 285-318.

Künzel, G., L. Korzun, and A. Meister. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics. 154(1):397-412.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, and J. Baldwin *et al*. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860-921.

Lindblad-Toh, K., C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, and M. Kamal *et al.* 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803-819.

Luce, A.C., A. Sharma, O.S. Mollere, T.K. Wolfgrober, K. Nagaki, J. Jiang, G.G. Presting, and R.K. Dawe. 2006. Precise centromere mapping using a combination of repeat junction markers and chromatin immunoprecipitation-polymerase chain reaction. Genetics. 174:1057-1061.

Mayer, K.F.X., R. Waugh, P. Langridge, T.J. Close, R.P. Wise, and A. Graner *et al.* 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature. 491:711-717.

Michalak de Jimenez, M.K., F.M. Bassi, F. Ghavami, K. Simons, R. Dizon, R.I. Seetan, L.M. Alnemer, A.M. Denton, M. Dogramaci, H. Simkova, J. Dolezel, K. Seth, M.C. Luo, J. Dvorak, Y.Q. Gu, and S.F. Kianian. 2013. A radiation hybrid map of chromosome 1D reveals synteny conservation at a wheat speciation locus. Funct. Integr. Genomics. 13:19-32.

Morell, M.K., B. Kosar-Hashemi, M. Cmiel, M.S. Samuel, P. Chandler, S. Rahman, A. Buleon, I.L. Batey, and Z. Li. 2003. Barley sex6 mutants lack starch synthesis IIa activity and contain a starch with novel properties. Plant J. 34:173-185.

Nevo, E., Y.B. Fu, T. Pavlicek, S. Khalifa, M. Tavasi, A. Beiles. 2012. Evolution of wild cereals during 28 years of global warming in Israel. Proc. Natl. Acad. Sci. USA. 109:3412-3415.

Nadeau, J.H. and W.N.B. Frankel. 2000. The roads from phenotypic variation to gene discovery mutagenesis versus QTLs. Nat. Genet. 25:381-384.

Nasuda, S., Y. Kikkawa, T. Ashida, A.K.M.R. Islam, K. Sato, and T.R. Endo. 2005. Chromosomal assignment and deletion mapping of barley EST markers. Genes. Genet. Syst. 80(5):357-366.

Ozturk, Z.N., V. Talame, M. Deyholos, C.B. Michalowski, D.W. Galbraith, N. Gozukirmizi, R. Tobeerosa, and H.J. Bohnert. 2002. Monitoring large-scale changes in transcript abundance in drought- and salt-stresssed barley. Plant Mol. Biol. 48:551-573.

Paux, E. D. Roger, E. Badaeva, G. Gay, M. Bernard, P. Sourdille, and C. Feuillet. 2006. Characterizing the composition and evolution of homoeologous in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J. 48:463-474.

Peters, J.L., F. Cnudde, and T. Gerats. 2003. Forward genetics and map-based cloning approaches. Trends Plant Sci. 8(10):484-491.

Poland, J., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping by sequencing approach. PloS ONE. 7(2):1-8.

Rostocks, N., K. Ramsay, K. MacKenzie, L. Cardle, P.R. Bhat, M.L. Roose, J.T. SAvensson, N. Stein, R.K. Varshney, D.F. Marshall, A. Graner, T.J. Close, and R. Waugh. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc. Natl. Acad. Sci. USA. 103(49):18656-18661.

Rudi, H., A.K. Uhlen, O.M. Harstad, and L. Munck. 2006. Genetic variability in cereal carbohydrate compositions and potentials for improving nutritional value. Anim. Feed Sci. Technol. 130:55-65.

Sabot, F., A.H. Schulman. 2009. Genomics of transposable elements in the Triticeae. In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p. 387-406.

Sakai, K., S. Nasuda, K. Sato, and T.R. Endo. 2009. Dissection of barley chromosome 3H in common wheat and a comparison of 3H physical and genetic maps. Genes Genet. Syst. 84:25-34.

Sakata, M., S. Nasuda, and T.R. Endo. 2010. Dissection of barley chromsome 4H in common wheat by the gametocidal system and cytological mapping of chromsome 4H with EST markers. Genes. Genet. Syst. 85:19-29.

Schulte, D., T.J. Close, A. Graner, P. Langridge, T. Matsumoto, G. Muehlbauer, K. Sato, A.H. Schulamn, R. Waugh, R.P. Wise, and N. Stein. 2009. The international barley sequencing consortium at the threshold of efficient access to the barley genome. Plant Physiol. 149:142-147.

Schulte, D., R. Ariyadasa, B. Shi, D. Fleury, C. Saski, M. Atkins, P. deJong, C.C. Wu, A. Graner, P. Langridge, and N. Stein. 2011. BAC library resources for map-based cloning and physical map construction in barley (*Hordeum vulgare* L.) BMC Genomics. 12:247-257.

Serizwa, N., S. Nasuda, F. Shi, T.R. Endo. 2001. Deletion-based physical mapping of barley chromosome 7H. Theor. Appl. Genet. 103:827-834.

Shi, F. and T.R. Endo. 1997. Production of wheat–barley disomic addition lines possessing an *Aegilops cylindrical* gametocidal chromosome. Genes. Genet. Syst. 72:243-248.

Shi, F., T.R. Endo. 1999. Genetic induction of structural changes in barley chromosomes added to common wheat by a gametocidal chromsomes derived from *Aegilops cylindrical*. Genes. Genet. Syst. 74:49-54.

Sreenivasulu, N., V. Radchuk, M. Strickert, O. Miresch, W. Weschke, and U. Wobus. 2006. Gene expression patterns reveal tissue-specific signaling networks controlling

programmed cell death and ABA-regulated maturation in developing barley seeds. The Plant J. 47(2):310-327.

Stein, N., M. Prasad, U. Scholz, T. Thiel, H. Zhang, M. Wolf, R. Kota, R.V. Varshney, D. Perovic, I. Gross, and A. Graner. 2007. A 1,000-loci transcript map of the barley genome:new anchoring points for integrative grass genomics. Theor. Appl. Genet. 114:823-839.

Stephens, J.L., S.E. Brown, N.L.V. Lapitan, and D.L. Knudson. 2004. Physical mapping of barley genes using an ultrasensitive flouorescence in situ hybridization technique. Genome. 47:179-189.

Stewart, E.A. and D.R. Cox. 1997. Radiation hybrid mapping. In: P. H. Dear, editor, Genome mapping. Oxford University Press, New York. p. 72-93.

Taketa, S., S. Amano, Y. Tsujino, T. Sato, D. Saisho, K. Kakeda, M. Nomura, T. Suzuki, T. Matsumoto, K. Sato. 2008. Barley grain with adhering hulls in controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. Proc. Natl. Acad. Sci. USA. 105:4062-4067.

Tanno, K., S. Taketa, K. Takeda, and T. Komatsuda. 2002. A DNA marker closely linked to the *vsr1* locus (two-type gene) indicates multiple origins of six-rowed cultivated barley (*Hordeum vulgare* L.) Theor. Appl. Genet. 104:54-60.

Timms, L., R. Jimenez, M. Chase, D. Lavelle, L. McHale, A. Kozik, Z. Lai, A. Heesacker, S. Knapp, L. Rieseberg, R. Michelmore, and R. Kesseli. 2006. Analyses of senteny between *Arabidoopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. Genetics. 173:2227-2235.

Turner, A., J. Beales, S. Faure, R. Dunford, and D. Laurie. 2005. The pseudoresponse regulator *Ppd-H1* provides adaptation to photoperiod in barley. Science. 310:1031-1034.

You, F.M., H. Wanjugi, N. Huo, G.R. Lazo, M.C. Luo, O.D. Anderson, J. Dvorak, and Y.Q. Gu. 2010. RJPrimers: unique transposable element insertion junction discovery and PCR primer design for marker development. Nucleic Acids Res. 38:313-320.

Deynze, A.E., J.C. Nelson, M.E. Sorrells, S.R. McCouch, J. Dubkovsky, J. Dvorak, K.S. Gill, B.S. Gill, E.S. Lagudah, and R. Apples. 1995. Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome. 38:45-59.

Vogel, J. and J. Bragg. 2009. Brachypodium distachyon, a new model for the Triticeae. In: C. Feuillet, G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer+Business media, LLC. p. 387-406.

Wade, C.M., E. Giulotto, S. Sigurdsson, M. Zoli, S. Gnerre, and F. imsland et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science. 6;326(5954):865-867.

Walter, M.A., D.J. Spillett, P. Thomas, J. Weissenbach, and P.N. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genome. Nature Genet. 7:22-28.

Wanjugi, H., D. Coleman-Derr, N. Huo, S.F. Kianian, M.C. Luo, J. Wu, O. Anderson, and Y.Q. Gu. 2009: Rapid development of PCR-based genome specific repetitive DNA junction markers in wheat. Genome. 52(6):576-587.

Waterson, R.H., K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, and P. Agarwal et al. 2002. Initial sequening and comparative analysis of the mouse genome. Nature. 5;420(6915):520-562.

Wenzl, P., H. Li, J. Carling, M. Zhou, H. Raman, and E. Paul et al. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics. 7:206-228.

Wi, G.S., B.Y. Chung, J.S. Kim, J.H. Kim, M.H. Baek, J.W. Lee, and Y.S. Kim. 2007. Effects of gamma irradiation on morphological changes and biological responses in plants. Micron. 38:553-564.

Zohary, D. 1969. The progenitors of wheat and barley in relation to domestication and agriculture dispersal in the old world. In: P.J. Ucko and G.W. Dimbley, editors, The domestication and exploitation of plants and animals. Duckworth, London. p. 47-66.

2. HIGH RESOLUTION RADIATION HYBRID MAP OF BARLEY CHROMOSOME 3H

Assembly of the barley genome is complicated by its large size (5.1 Gb) and highly repetitive nature (84%). This process is facilitated by high resolution maps for aligning available BAC contigs along chromosomes. Available genetic maps do not provide accurate information on the physical position of a large portion of the genome located in recombination-poor regions. Radiation hybrid (RH) mapping is an alternative approach, which is based on radiation-induced deletions along chromosomes. In recent decades, RH maps have replaced conventional genetic maps in human and animal genomes due to their even coverage of all chromosomal regions and higher resolution. Here, we developed 373 *in vivo* RH lines of barley chromosome 3H. The RH population was generated by irradiating wheat-barley chromosome 3H addition lines and crossing them to a normal wheat cultivar. The average induced deletion frequency of informative RH lines was 8.56%, with each line containing on average three deletions of 294 Kb in length. The induced deletion size varied from 281 bp-4.58 Mb. Compared to an analogous genetic map, an initial 3H-RH map had a total of 9.53-X higher resolution, reaching a maximum of >262.40X in regions around the centromere. The final RH map consisted of 113 markers spanning 2,577.4 cR in length, with an average resolution of 2.22 Kb. This resolution level enabled alignment of randomly picked BAC contigs ranging from 1.50 Kb to 22.95 Kb in size. Induced deletions were not randomly distributed along the chromosome, and some regions, such as the short arm, showed reduced deletion frequency. The uneven deletion distribution affected the map uniformity, with a maximum 8.91-fold deviation from the average map resolution. The 3H-RH map developed in this study is the first barley RH map and has the highest resolution among plant RH maps. The high resolution and the coverage of poor-recombination regions make the

generated RH map an ideal resource for barley genome assembly, as well as other genetic studies.

2.1. Introduction

Barley (*Hordeum vulgare* L., 2n=2X=28) is ranked fourth in production among cereals and is cultivated on about 50 million hectares worldwide (FAO, 2012; http://faostat.fao.org/). Compared to other cereals, barley is more tolerant to drought and salinity stress. Therefore, it is a main food source in countries with harsh environmental conditions (Nevo et al., 2012). Due to the diploid genome and the self-pollinated characteristic, barley is a model plant for genetic studies of Triticeae crops with complex genomes such as wheat (*Triticum aestivum* L., *T. turgidum* L.) and rye (*Secale cereale* L.) (Gaut, 2002). A complete reference of the barley genome sequence would reveal genetic networks underlying important traits and enable exploitation of the full potential of barley and related Triticeae crops in breeding programs. Despite concerted international efforts, complexity of the genome (5.1 Gb in size with 84% consisting of repeat sequences [Mayer et al., 2012]), preclude the complete genome assembly based on existing resources.

The current approach for barley genome assembly is to use high resolution genetic maps for ordering available BAC contigs (Mayer et al., 2012). To date numerous barley genetic maps have been developed (http://wheat.pw.usda.gov/ggpages/map_summary.html). Genetic maps are based on recombination between polymorphic DNA sequences, represented as molecular markers. A limitation with genetic maps is the variation in recombination rates along chromosomes; recombination frequency is almost zero in the centromeric regions and increases by approximately square of the relative distance from the centromere (Kunzel et al., 2000; Akhunov et al., 2003). Thus, genetic maps do not reflect the actual physical distances of mapped

loci. Substantial variation of physical/genetic distance ratio is evident in the barley genome, ranging from 1.5 Mb/cM on the distal region to 89 Mb/cM on the pericentromeric region of chromosomes (Stephens et al., 2004). Moreover, one third of the Triticeae genome, located in recombination-poor regions (Erayman et al., 2004; Akhunov et al., 2003), cannot be mapped. Attempts to highly saturate barley genetic maps could not separate markers around the centromeric region (Wenzl et al., 2006). Due to this limitation, 1.2 Gb of barley contigs were not aligned along chromosomes even with the use of saturated genetic map platform with half-million markers (Mayer et al., 2012).

To better estimate the physical locations of loci, deletion-bin maps, also known as gametocidal maps, were generated. Introgression of an alien chromosome from *Aegilops cylindrical* to bread wheat containing a barley addition chromosome induces terminal deletions on the addition barley chromosome (Endo 2007). With assigning markers to induced deletion bins, deletion-bin maps of some barley chromosomes were generated (Serizawa et al., 2001; Masoudi-Nejad et al., 2005; Ashida et al., 2007; Sakai et al., 2009; Sakata et al., 2010; Joshi et al., 2011). The drawback to deletion-bin maps is that they do not separate markers located at distal chromosomal regions (Sakai et al., 2009), and the order loci within single bins cannot be determined.

An alternative approach is Radiation hybrid (RH) mapping. In this method, ionizing radiation (e.g. gamma-ray) is used to induce deletions across chromosomal length. The chromosome(s) of interest is then rescued by integrating into non-native recipient cells. The frequency of induced breakages between two loci is used as a distance measure; markers that are physically close show few breakages compared to more distant markers. The frequency of breakages between all pairwise combinations of loci in an RH population is calculated to

construct RH maps. The RH map unit is defined as a centi-Ray (cR), corresponding to one break between two loci in every 100 lines (Womack et al., 1997). RH approach has three main advantages compared to genetic maps: *I)* RH maps cover the whole chromosomal regions; in other words, RH maps have even uniformity; *II)* RH map resolution can be improved by inducing more deletions through increasing the radiation dosage (Kumar et al., 2014); and *III)* both polymorphic and monomorphic markers can be applied to genotype an RH population.

The RH approach was first described by Goss and Harris (1975) by fusing irradiated human X chromosomes with rodent cells. Since then, RH method has been adapted to develop numerous maps of human (Hudson et al., 1995; Schuler et al., 1996; Stewart et al., 1997) and animal chromosomes (for review see Faraut et al. 2009). Due to the high resolution and uniformity, generated RH maps were successfully used for genome assembly in several sequencing projects (Lander et al., 2001; Gibbs et al., 2004; Lindblad-Toh et al*.,* 2005). In plants, available cytogenetic stocks of addition and substitution lines can accelerate RH population development by replacing the process of *in vitro* cell fusions. Using theses cytogenetic stocks, several RH maps were developed for maize chromosome 1 (*Zea mays* L.) (Kynast et al., 2004), bread wheat chromosome 1D and 3B (Kalavacharla et al., 2006; Kumar et al., 2012a), and *Aegilops tauschii* chromosome 2D (Kumar et al., 2012b). In barley, a whole-genome RH panel was developed by fusing irradiated barley (*H. vulgare*) protoplast with tobacco (*Nicotiana tabacum* L.) cells. Using this technique 40 *in vitro* RH lines were developed (Wardrop et al., 2002); however, due to the insufficient DNA quantity, generated lines were not genotyped for mapping the barley genome.

Despite the extensive application of RH technique in mapping animal genomes, this technique is applied in few plant genomic studies, and the majority of plant RH maps are in

preliminary stages. At the time of this study, the RH map of bread wheat chromosome 3B was the only high resolution continuous RH map in plants, spanning 1871.9 cR, with 90 Kb resolution (Kumar et al., 2012a). Generated RH map had 10-fold higher resolution and was six-fold more uniformed than a similar quality genetic map (Kumar et al., 2012). The 3B-RH map was used as a platform in genome assembly of the chromosome 3B (Paux et al., 2008).

In this study, an *in vivo* RH population, consisting of 373 lines, was developed for barley chromosome 3H. After constructing an RH map, the map resolution across different regions was determined. Mapping efficiency was then compared between the RH map and an analogous genetic map. Finally, utility of the RH map for genome assembly was confirmed by anchoring randomly picked contigs to the generated map.

2.2. Materials and Methods

2.2.1. RH population development

Addition lines of bread wheat cultivar "Chinese Spring" (CS), disomic for barley (*H. vulgare* L.) cultivar "Betzes" chromosome 3H (CS+3H"), developed by Islam et al. (1981), were used. The addition line seeds were tempered via atmospheric hydration by placing them in an airtight vessel between open-top jars containing an aqueous glycerol solution (60% v/v) for ten days, as described by Hossain et al. (2004). Tempered seeds were irradiated with gamma-rays at three different levels: 15, 25, and 35 Krad, using 150 seeds for each radiation treatment. Irradiated seeds were treated with Raxil® MD fungicide (0.1X) and germinated in petri-dishes on moist filter paper. Germinating seeds were kept in a dark cold room (4° C) for two days, after which they were placed in normal room conditions (23° C) for 14 days. Germinated seeds (RH_0 lines) were planted in seedling trays containing Sunshine Mix#1/LC1 and transplanted in a greenhouse soil bed after 21days. Survival rate was calculated as the percentage of germinated

lines that survived under greenhouse conditions compared to 100% survival rate of control lines with no radiation treatment. From the survived RH0 population (CS+3H"), 94 lines (47 lines from 15 and 25 Krad radiation treatments) were selected for crossing. One to three heads of each RH_0 line were crossed, as female, with CS (male parent). Resulting seeds were planted in a greenhouse and constituted the RH population. At each generation (RH, RH_0, and original wheat-barley addition seed stocks), presence of the 3H addition chromosome was confirmed with ten 3H chromosome-specific molecular markers distributed along the chromosome. These markers included five expressed sequence tags (ESTs) selected from a deletion bin map (Sakai et al., 2009) and five repeat junction markers (RJMs) designed in this study, based on the sequences of chromosome 3H contigs with known genetic positions (Sato et al., 2011).

2.2.2. Molecular markers design and PCR analysis

To analyze radiation-induced chromosome 3H deletions, 93 ESTs previously assigned to chromosome 3H using wheat-barley addition lines (Nasuda et al., 2005) were used. In addition, 72 genome-specific RJMs were designed using "RJPrimer" software (http://probes.pw.usda.gov/RJPrimers/) from chromosome 3H contig sequences obtained from sequences installed on the Gbrowse system (http://150.46.168.145/gbrowse/). One wheat-specific molecular marker was applied in multiplex PCR reactions along with each chromosome 3H specific marker to provide a positive control against PCR miss-amplifications. DNA material was extracted at the four-leaf stage as described by Guident et al. (1991). PCR was carried out in 20 µl reactions containing 30 ng of DNA, 1X PCR buffer, 1.5 mM of $MgCl_2$, 250 µM of each dNTP, 250 µM of each primer, and 1.0 unit of *Taq* DNA polymerase. Touchdown PCR was performed as follows: 94°C for 4 min.; five cycles of 94°C for 30 sec., 65°C for 30 sec., 72°C for 1 min., with the annealing temperature decreased 1°C per cycle; 35 cycles of 94°C for 30 sec.,

60°C for 30 sec., and 72°C for 1 min., followed by 72°C for 7 min. PCR products were visualized using agarose (1.5% SeaKem® LE Agarose) gel electrophoresis, 100V for 50 min.

2.2.3. Characterizing the RH population

A total of 328 3H-RH lines were genotyped with 113 markers. RH lines were scored for the presence or absence of markers. Ambiguous bands were scored as missing data. Deletion frequency within each line was estimated by calculating the percentage of number of deleted markers. The retention frequency of each line was estimated by subtracting the deletion frequency of that line from 100. To characterize irradiation–induced deletions and to generate an RH map, genotyping data of RH lines with 0<retention frequency<100 were used. The deletion frequency of each marker was estimated by calculating marker deletion percentage across the population. The size of each deletion was estimated based on the physical positions of markers flanking that deletion. In the absence of markers with known physical positions, deletion size was estimated based on the RH positions of markers flanking deletions and converting the deletion RH distance to physical distance.

2.2.4. Developing the RH map

The RH map was generated by Carthagene 1.2.2 (de Givry et al., 2005) with the minimum LOD of 2.0. Thirty-three ESTs with known physical and/or genetic positions (Sakai et al., 2009) were first mapped to generate a skeleton RH map. Remaining markers were then added to the anchored marker intervals using the buildfw function. This command developed all possible marker combinations within each interval and selected the best interval for each marker, based on the highest LOD score. Markers assigned to intervals were then mapped by the same command. The generated map was improved using the greedy search, flips, genetic algorithm, annealing, and polish functions. Data from an additional 45 RH lines genotyped with 39 markers

were integrated to the generated map to improve the developed RH map, using the dsmergen command.

2.3. Results

2.3.1. 3H-RH population characterization

An *in vivo* 3H-RH population was developed by gamma-irradiating CS+3H" seeds and then crossing the irradiated plants to normal CS. Three levels of gamma-ray (15, 25, and 35 Krad) were used to determine the optimum radiation dosage that induces maximum deletions without severely damaging seed germination and plant development. Compared to 100% survival with no radiation treatment, the survival rate decreased to 96.93% and 85.85% at 15 Krad and 25 Krad, respectively. The germination rate dropped to 0.66% at 35 Krad (Figure 2.1). In the RH_0 generation (CS+3H"), the induced chromosome deletions might be masked by homologous chromosomes. Thus, RH_0 lines were crossed to CS to separate chromosome 3H homologs. About half of the crossed progenies, totaling 373 seeds, germinated and constituted the 3H-RH populations (CS+3H').

For genotyping the RH population, chromosome 3H-specific markers were required so that they would not be interfered by the syntenous wheat background. With this in mind, we used EST markers specific for chromosome 3H (Nasuda et al., 2005) and genome-specific RJMs, as described in Materials and Methods. A Total of 164 markers were tested on the barley cultivar "Betzes" (positive control) and wheat cultivar "CS" (negative control) in PCR amplifications. In total 113 markers produced amplicons from "Betzes" without amplifying any DNA fragments from "CS". These markers were selected to characterize the RH population and generate 3H-RH map. Selected markers consisted of 65 ESTs and 48 RJMs, among which 41 RJMs were

designed from 22 contigs distributed across a genetic map of the chromosome 3H (Sato et al., 2011).

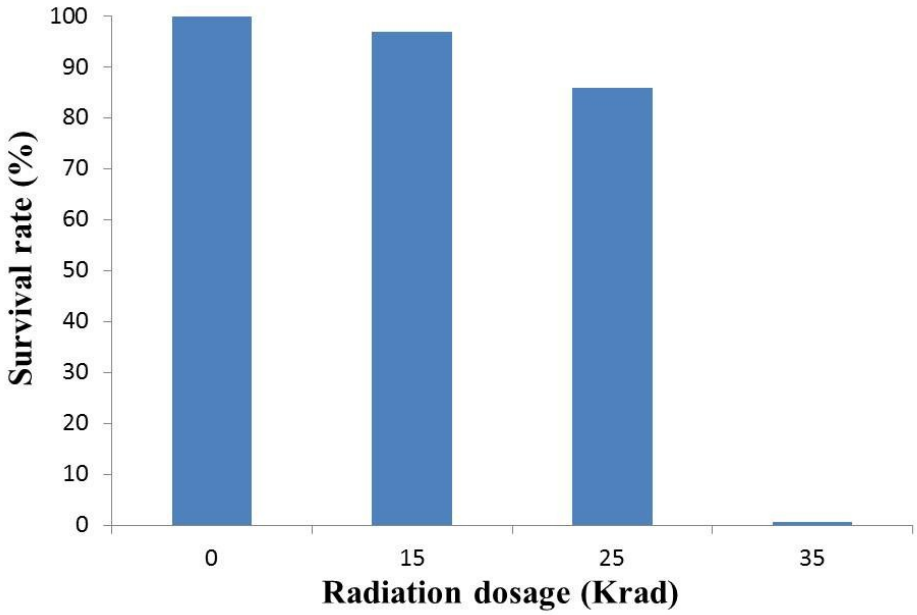

Figure 2.1. Survival rate of CS+3H" addition lines treated with the gamma-radiation
Groups of 150 seeds were irradiated with three radiation dosages (15, 25, and 35 Krad). Survival rate was calculated based on the 100% survival rate of the control treatment (0 radiation dosage).

The retention frequency value indicates the proportion of a chromosome that was retained after irradiation. Retention frequency is presented on a scale from zero to one, with zero representing no deletions and one representing complete elimination of the chromosome. Genotypic data of 113 selected markers showed that 27.13% of the RH lines lost the whole chromosome 3H, whereas 7.31% of the lines had no deletions on the chromosome (Figure 2.2). Loss of the whole chromosome 3H might be due to the deletion of a critical chromosomal region, such as the centromere, which prevented proper chromosome segregation. RH lines with zero and one retention frequency were excluded from further analysis for they do not provide any mapping information. The remaining 215 lines were designated as informative lines.

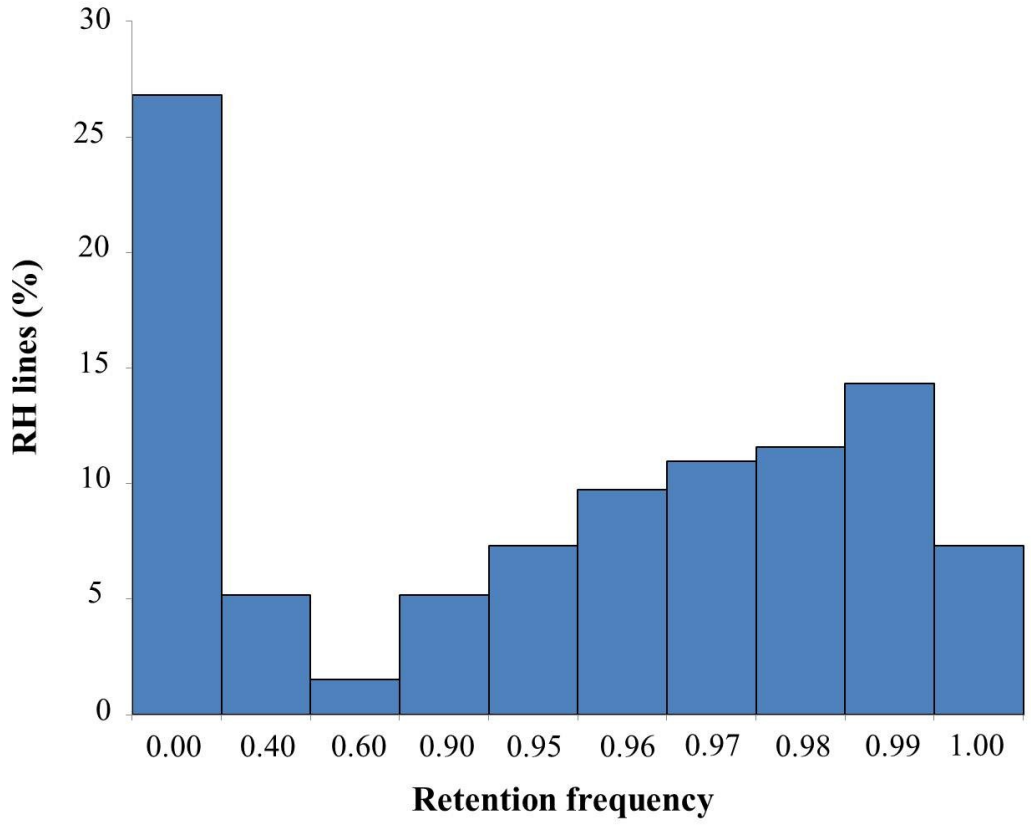

Figure 2.2. Distribution of the chromosome 3H retention frequency among 3H-RH lines
Retention frequency was calculated for 240 lines genotyped by 113 markers.

The average chromosome 3H-retention frequency of informative lines was 0.91, ranging from 0.03-0.99. The majority of RH population (59.11% of the RH population or 89.81% of informative lines) had the retention frequency of 0.90-0.99 (Figure 2.2). Lines with 0.9 retention frequency composed 5.18% of the population, this percentage increased gradually to 14.32% in lines with 0.99 retention frequency. The retention frequency pattern observed in this study was similar to what was reported in bread wheat chromosome 3B-RH population (Kumar et al., 2012a).

A single deletion was considered as the deletion of one marker or continuous deletion of markers located next to one another. Based on this assumption, each line had an average of three deletions, ranging from one to twelve deletions, across the chromosome 3H. The average

deletion size was 292.88 Kb, varying widely from 281 bp to 4.57 Mb. The largest deletion spanned from the mid region of the short chromosome arm (RJH2477/2 marker at 545.7 cR) to the long arm telomeric region (RJH856/24 marker at 2604.3 cR).

Deletion frequency (deletion percentage of a given marker across the RH population) was calculated to determine if deletion inducement differs among chromosomal regions. The average deletion frequency among informative lines was 8.56%, which was close to what was reported in bread wheat (Kumar et al., 2012a). ESTs detected 10.20% and RJMs 6.34% of the deletions. Deletion frequency was different among chromosomal regions (Figure 2.3). In general, the long arm deletion frequency (11.89%) was 3.37-fold higher than the short arm (3.52%). The maximum deletion frequency was observed in three segments on the long arm: two segments in the peritelomeric region with 25.47% and 35.50% deletion frequency and one segment in the pericentromeric region with 25.73% deletion frequency. Surprisingly, increasing radiation dosage from 15Krad to 25Krad decreased the deletion frequency by 3.20% (from 10.26% deletion frequency at the 15 Krad dosage to 7.06% at the 25 Krad dosage). Higher radiation dosage, however, increased the number of lines with deletion(s). Among lines with no deletions, 29.16% were irradiated with 25 Krad dosage and 70.83% with 15 Krad dosage. Also, the 25 Krad dosage induced more small deletions compared to the 15 Krad dosage. Thus, one could conclude that high radiation dosage induces more frequent but smaller deletions, which due to their small size, remained undetected. It is expected that by using more markers, additional 25 Krad-induced deletions would be detected, and the deletion frequency induced by the 25 Krad dosage would increase to a higher level than those induced by the 15 Krad dosage.

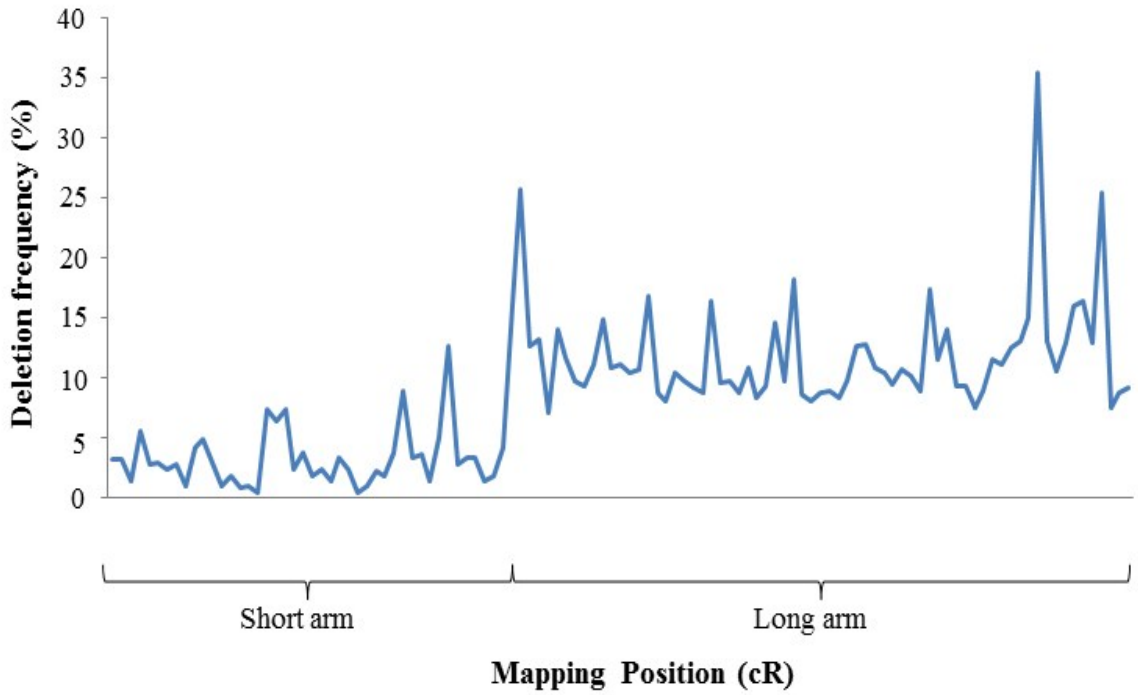

Figure 2.3. Distribution of radiation-induced deletion frequency along the barley chromosome 3H Deletion frequency was calculated for 215 informative lines genotyped with 113 markers.

2.3.2. Comparing the 3H-RH map with genetic and deletion bin maps

Genotypic data from 33 ESTs previously mapped on genetic and deletion bin maps (Sakai et al., 2009) were used to develop an initial skeleton RH map. Map efficiency of the initial RH map was compared with that of the genetic and deletion-bin maps. The skeleton RH map was 1,530.2 cR in length and the marker order was consistent with both genetic and deletion bin maps. The RH map separated all markers that were clustered on previous maps (Figure 2.4). For example, four markers (k03504-k03692) that were mapped at one locus on centromeric region of the genetic map occupied an interval of 389.1 cR (from 280.4 cR to 669.5 cR) on the RH map. Two markers (k0892 and k03336) flanking the centromeric region on the deletion bin map were used to distinguish chromosome arms on the RH and genetic maps. Those two markers located the centromeric region in a 200.5 cR interval of the RH map. Marker clusters were not

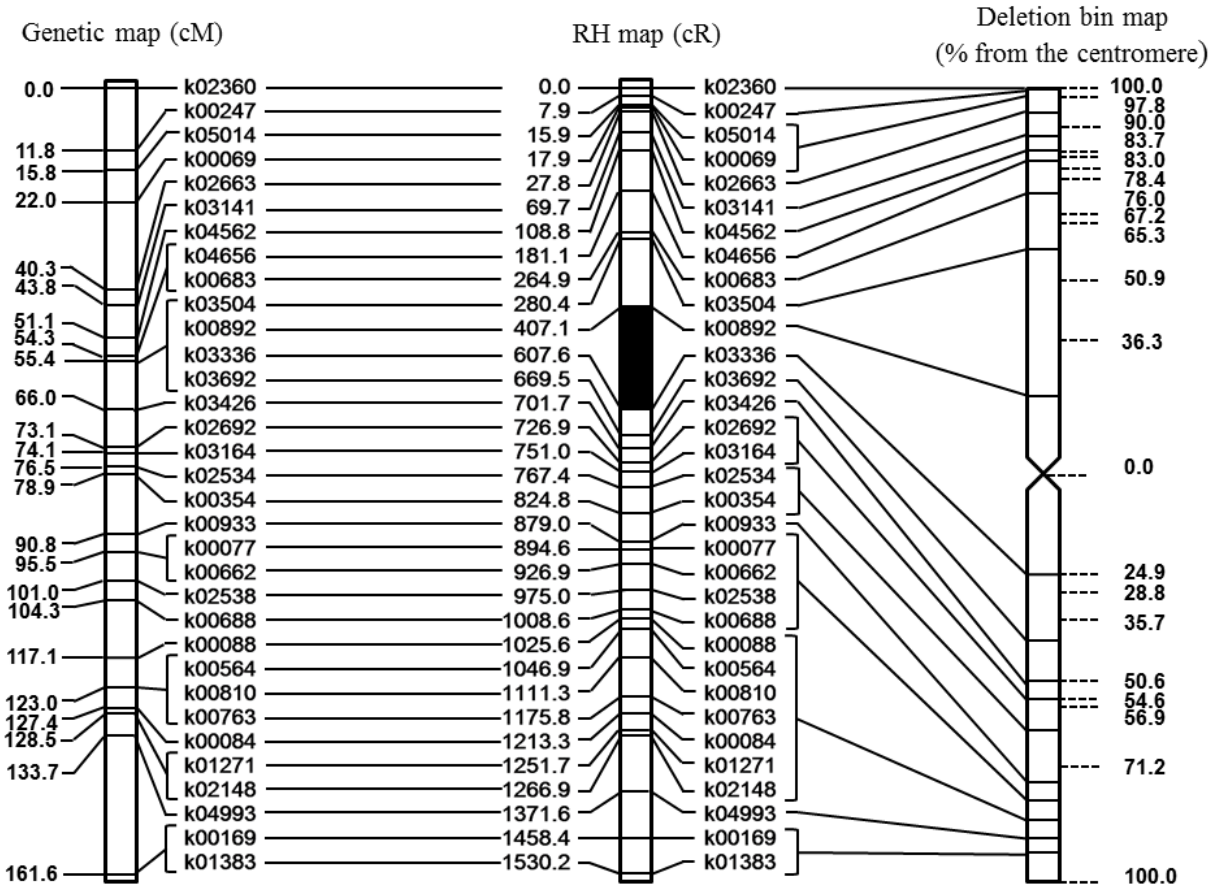

Figure 2.4. Skeleton RH, genetic, and deletion bin maps of the barley chromosome 3H The same markers on chromosome 3H genetic and deletion bin maps (Sakai et al., 2009), were used to develop the skeleton RH map. The RH map centromeric region is shown in black.

limited to the genetic map centromeric region. The majority of clusters were found on the long arm of both genetic and deletion bin maps. For example, seven markers (k00088-k02148) were located at one locus on the long arm distal region of the deletion bin map, which were mapped in a 241.3 cR interval (from 1025.6 cR to 1266.9 cR) on the RH map.

Map resolution of the skeleton RH map was compared to the analogous genetic map. Map resolution is defined as the minimum physical distance between two loci required for mapping them in separate positions (Kumar et al., 2012). It is estimated by dividing the physical length of a given region by the map length of that region, with a smaller ratio indicating higher resolution. Considering that chromosome 3H is 755 Mb in length (Mayer et al., 2011), the RH

map resolution was 0.49 Mb (755 Mb/1,530.2 cR), which was 9.53-fold higher than the 4.67 Mb (755 Mb/161.6 cM) genetic map resolution.

Resolution of the genetic and RH maps was compared in greater detail at different chromosomal regions by comparing the intervals of the 33 mapped markers. Considering that the same markers identify the same loci, map resolution of both genetic and RH maps can be compared. The ratio of RH distance/genetic distance was used as a comparative indicator of resolution. This ratio was calculated for different chromosomal regions. We used physical positions of the markers mapped on the deletion-bin map (Figure 2.4) to divide the chromosome into five chromosomal regions: short arm telomeric, short arm middle, pericentromeric, long arm middle and long arm telomeric regions (Table 2.1). In the pericentromeric region, the RH map resolution had a maximum of >262.40-fold higher resolution than the genetic (Table 2.1). After pericentromeric region, the highest RH distance/genetic distance value was 205.45 in the short arm middle region. In contrast, RH distance/genetic distance value of the long arm middle region was 5.61, which was considerably lower than the ratio in the corresponding region on the short arm. The RH distance/genetic distance value was 3.34 and 9.62 in the telomeric region of the short and long arm, respectively.

It should be considered that in this study we compared an RH map with analogues genetic and deletion-bin maps, despite the fact that mapping populations were derived from different barley varieties. Recombinant population was derived from a cross between barley var. Haruna Nijo and accession H602, whereas RH and deletion-bin populations were developed based on barley var. Betzes. Also, all the three maps were comprised of ESTs, which are concentrated at distal regions of Triticeae chromosomes (Lehmensiek et al., 2009). Using same

mapping population and large number of markers, distributed across all chromosomal regions would result in a more precise comparison.

Table 2.1. Comparison of the skeleton RH map with the analogous genetic map of the barley chromosome 3H The chromosome 3H was divided into five regions based on the marker physical positions from the centromer (Sakai et al., 2009). The genetic position of marker were obtained from the chromosome 3H genetic map developed by Sakai et al. (2009).

	Chromosomal regions				
	Short arm telomeric	**Short arm middle**	**Pericentromeric**	**Long arm middle**	**Long arm telomeric**
Position from centromere (%)	80-100	80-20	20 from short arm-25 from long arm	25-80	80-100
No. of markers	7	3	2	7	14
Genetic map length (cM)	54.3	1.1	0	40.1	66.1
RH map length (cR)	181.1	226	262.4	225.1	635.6
RH length/genetic length	3.34	205.45	>262.40	5.61	9.62

2.3.3. Developing a comprehensive chromosome 3H-RH map

Genotypic data of an additional 80 markers were integrated into the skeleton RH map. The final map consisted of 113 markers covering over a 2,577.4 cR region, with the average LOD of 17.76 (Figure 2.5). Four markers (RJH15851/9, RJH608/5, k03410, RJH747/1) were integrated within an interval of the pericentromeric region (k03504-k03692 interval), which were not separated on the genetic map (Fig. 2.4). None of markers used in this study were mapped within the centromeric region (k00892-k03336 interval). This result was expected for we used gene-based markers (ESTs and RJMs bearing ESTs), which have a low density in the centromeric region (Lehmensiek et al., 2009).

The final map resolution was estimated based on the resolution of six BAC contigs (with an average length of 40,179 bp), anchored on the RH map (Table 2.2, Figure 2.5). These contigs were selected based on their even distribution along the chromosome; three evenly spaced contigs were selected to represent different chromosomal regions: peritelomeric,

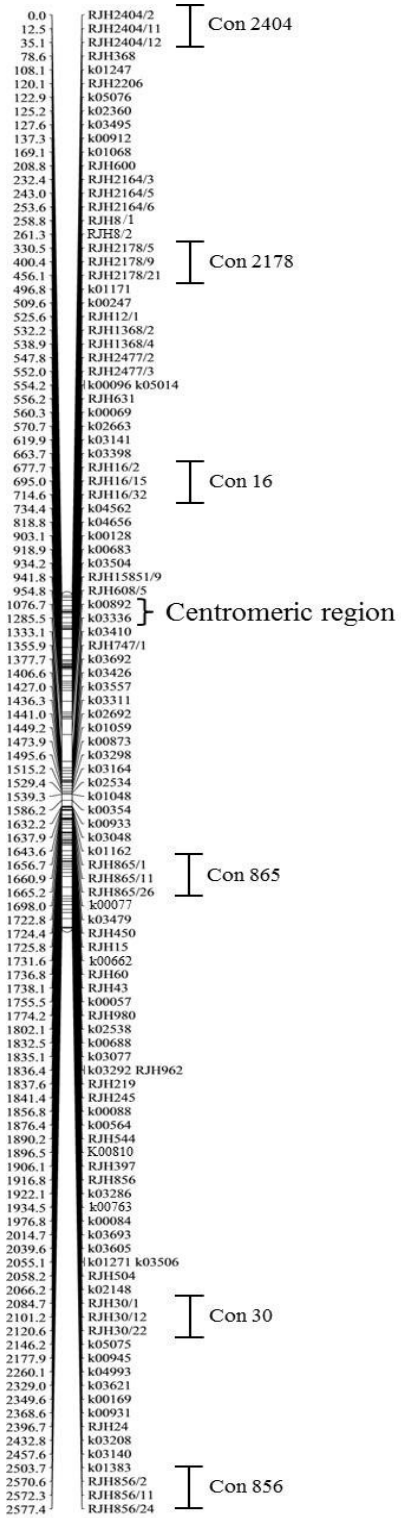

Figure 2.5. The comprehensive RH map of barley chromosome 3H Final map consisted of 113 markers, extending over a 2577.4 cR distance. Contig-based markers that were used for calculating map resolution are marked with error bars. Markers designed from single contigs are mapped next to one other and are distinguished by a slash (/).

Table 2.2. RH positions of BAC contigs representing six regions of barley chromosome 3H
Markers k03336 and k00892, flanked the centromeric region on a deletion bin map (Sakai et al., 2009), distinguished between chromosome arms

Representative. contig	Corresponding chromosomal region	Flanking markers	Region interval[1] (cR)	Interval length (cR)
2404	Short arm peritelomere	RJH2404/2-RJH2178/5	0-329	329
2178	Short arm mid. region	RJH2178/5-RJH16/2	329-678.6	349.6
16	Short arm pericentromere	RJH16/2-k00892	678.6-1104.4	425.8
865	Long arm pericentromere	k03336-RJH865/1	1313.2-1684.1	370.9
30	Long arm mid. region	RJH865/1-RJH30/1	1684.1-2111.9	427.8
856	Long arm peritelomere	RJH30/1-RJH856/2	2111.9-2597.6	485.7

[1] RH position of flanking markers designed from the beginning sequence of the representative contigs were used to calculate the RH distance between the contigs.

pericentromeric, and the region in between, which we referred to as the middle (mid) region (Table 2.2). For a better estimation of map resolution, each contig was divided into two intervals by designing markers from the beginning, middle, and end sequences. Within each contig, the map resolution was calculated by taking the average resolution of its two intervals (Table2.3), and the final map resolution was calculated by taking the average resolution of all six contigs. Based on this calculation, the total map resolution was 2,228 bp with the maximum of 250 bp and minimum of 9,137 bp (8.91- and 4.10-fold deviation from the average, respectively) (Table 2.3). Overall, the short chromosome arm had 6.3-fold higher resolution (average 606 bp) compared to the long arm (average 3,849 bp). The middle region of each chromosome arm had the highest resolution (250 bp on the short arm and 1,031 bp on the long arm). The next highest resolution region was the pericentromere (716 bp on the short arm and 1,380 bp on the long arm). The peritelomeric region had the lowest resolution (819 bp on the short arm and 9,137 on the long arm). It should be considered that map resolution in this study was assessed based on six

selected contigs. Testing more contigs would provide a more definitive value of the map resolution.

Table 2.3. The RH map resolution across different chromosomal regions Map resolution (bp/cR) of each contig was calculated by taking the average resolution of its two intervals.

Chromosomal region	Designed markers	Map position (cR)	Marker intervals	Physical interval (bp)	RH interval (cR)	bp/cR	Average bp/cR
Short arm peritelomeric	2404/2 2404/11 2404/22	0.0 12.5 35.1	RJH2404/2-RJH2404/11	10,097	12.4	807.76	819.32
			RJH2404/11-RJH2404/22	18,778	22.6	830.88	
Short arm mid. region	2178/5 2178/9 2178/21	330.5 400.4 456.1	RJH2178/5-RJH2178/9	11,204	69.7	160.28	249.83
			RJH2178/9-RJH2178/21	18,904	55.5	339.38	
Short arm pericentromeric	16/2 16/15 16/32	677.7 695.0 714.6	RJH16/2-RJH 16/15	17,986	18.1	1,039.65	749.49
			RJH 16/15-RJH 16/32	9,003	20.5	459.33	
Short arm							606.21
Long arm peritelomeric	865/2 865/11 865/26	1656.7 1660.9 1665.2	RJH865/2-RJH865/11	5,952	4.2	1,417.14	1380.31
			RJH865/11-RJH865/26	5,777	4.3	1,343.48	
Long arm mid. region	30/1 30/12 30/22	2084.7 2101.2 2120.6	RJH30/1-RJH30/12	19,482	16.4	1,180.72	1031.15
			RJH30/12-RJH30/22	17,103	19.4	881.59	
Longt arm pericentromeric	856/2 856/2 856/2	2570.6 2572.3 2577.4	RJH856/2-RJH856/11	27,498	1.7	16,175.29	9136.95
			RJH856/11-RJH856/24	10,703	5.0	2,098.62	
Long arm							3849.47
Overall Chromosome							2227.84

Theoretically, the average 2.22 Kb resolution of the generated map enables anchoring contigs with a minimum length of 2.22 Kb onto the map. To confirm this hypothesis three random contigs with unknown mapping position were picked from contig viewer (Table 2.4). Using the generated RH map, we were able to order one contig as small as 1,50 Kb. The two other contigs, with 10.22 and 22.95 Kb lengths, were also ordered. These examples confirm the efficacy of the generated RH map in placing contigs onto the chromosome 3H.

Table 2.4. Alignment of chromosome 3H contigs to the RH map

Contig name	Contig size (kb)	Anchored markers	Map position (cR)
Con2164	1.50	RJH2164-3 RJH2164-5 RJH2164-6	232.4 243.0 253.6
Con1368	22.95	RJH1368-2 RJH1368-4	532.2 538.9
Con2477	10.22	RJH2477-2 RJH2477-3	547.8 552.0

2.4. Discussion

2.4.1. RH maps in assembly of the barley genome

In this study, the first RH map of a barley chromosome was generated. Comparing the initial RH map to a genetic map with the same quality showed that the RH map had an overall 9.53-X higher resolution, reaching to >262.40-X better resolution around the centromeric region (Table 2.1). This result was comparable with the wheat-RH map resolution, which was overall 10.5-fold higher than a genetic map, with a maximum of 136-X higher resolution in the centromere (Kumar et al., 2012a).

The final 3H-RH map consisted of 113 markers spanning over 2577.4 cR region. After the RH map of wheat chromosome 3B (Kumar et al., 2012a), this map was the second longest continuous RH map reported in plants. Two evidence validated the accuracy of the generated map: First, the RH map marker order was in total agreement with the marker order of genetic and deletion bin maps developed by Sakai et al., 2009 (Figure 2.4). Second, markers designed from single contigs (average length of 18,717 bp) were located next to one another on the RH map (Figure 2.5). The final map resolution was ~2.22 Kb, which was 40-fold higher than the 90 Kb resolution of wheat RH map (Kumar et al., 2012a). Induced deletions as small as 281 bp were identified in the 3H-RH population, which implied numerous small deletions might remain

undetected. Increasing the number of mapped markers most likely would increase the detection of these deletions and consequently would increase the map resolution.

The generated RH map enabled mapping loci located in pericentromeric region. This map, however, did not cover the centromeric region (~23% of the chromosome length around the centromere) because at the time of this study we lacked access to centromeric sequences for marker design. Many RH studies on plants and animals, however, have confirmed the existence of radiation-induced deletions within the centromeric region (Cox et al., 1990; Stewart et al., 1997; Riera-Lizarazu et al., 2000; Kurar et al., 2003; Pitel et al., 2004; Kalavacharla et al., 2006; Kumar et al., 2012 a). Therefore, we expect that integrating centromeric markers into the 3H-RH map, would saturate the centromeric region.

The 3H-RH map, due to the high resolution (2.22 Kb) and the high coverage, can bridge the resolution gap between genetic and contig-based physical maps. Using the generated RH map enabled anchoring randomly selected contigs, as small as 1.50 Kb, onto chromosome 3H. The 3H-RH map can serve as a template for developing RH maps of the remaining barley chromosomes. Barley RH maps would be a great platform for assembly of the barley genome and filling the 1.2 Gb gap of non-assembled contigs (Mayer et al., 2012) within the genome sequences.

2.4.2. Uneven deletion frequency along the chromosome

Preliminary analysis of plant RH panels suggested that induction of radiation-mediated deletions might be independent of chromosomal regions (Riera-Lizarazu et al., 2000; Kalavacharla et al., 2006; Kumar et al., 2012a). Our data, however, showed that induced deletions are not evenly distributed along the chromosome (Figure 2.3). For example, the deletion frequency on chromosome 3H short arm (3.52%) was 3.37-fold less than on the long

arm (11.89%). This variation was mainly due to 8.79% of 3H-RH lines having large induced deletions on the long arm, extending over >91.21% of the long arm, whereas the majority of short arm deletions were smaller in size (data not shown). A similar deletion pattern was observed in chicken chromosome 5 in which the short arm deletion rate was 15% less than that of the long arm (Pitel et al., 2004). The unequal deletion frequency was also reported in other animal genomes (Stewart et al., 1997; Kurar et al., 2003; Pitel et al., 2004) and wheat chromosome 1D and 7D (Kumar et al., 2012b).

Compared to a 13% deletion frequency along the wheat chromosome 3B (Kumar et al., 2012a), the short arm of barley chromosome 3H had an unusual resistance to induced deletions. To explain this observation, the short arm might be less condensed than the long arm. After inducing deletions, a DNA repair complex bridges two broken ends to fix the DNA breakage (Bleuyard et al., 2006). It is hypothesized that the repair complex requires open chromatin regions to function, which leads to small induced deletions on open chromatin regions and large deletions on more compressed regions (Kumar et al., 2012a). Assuming this hypothesis is true, the low deletion frequency of the short arm might suggest that the short arm is less condensed than the long arm.

The unequal deletion frequency led to an uneven map resolution along the chromosome. The final 3H-RH map had a maximum of 8.91-fold deviation from the total map resolution (Table 2.3), which was similar to the five-fold deviation from the resolution of wheat chromosome 3B-RH map (Kumar et al., 2012a). Despite the 8.91 resolution deviation within 3H-RH map, this map was relatively uniform compared to genetic maps with large resolution fluctuations (e.g. up to 60-fold in a barley genetic map) (Stephen et al., 2004).

The uneven RH map resolution was also evident on the long arm of the skeleton 3H-RH map. It is known that the genetic map resolution increases from the centromere toward telomers (Kunzel et al., 2000; Akhunov et al., 2003; Kumar et al). If the RH map resolution was uniform, the ratio of RH map resolution to genetic map resolution (RH distance/ genetic distance) would decrease from centromeic region toward telomeres. This was the case on the short arm, which RH distance/ genetic distance ratio decreased from >262.40 in the near centromeric region to 225 and 3.34 in the middle and the peritelomeric region, respectively. In the long arm; however, this ratio dropped from 262.40 in the near centromeric region to 5.61 in the middle region, after which increased to 9.62. The fluctuation in RH distance/ genetic distance ratio on the long arm could be explained by the uneven distribution of irradiation-induced deletions along the chromosome arm (Figure 2.3).

2.4.3. Factors involved in improving RH mapping

2.4.3.1. Deletion pattern

RH mapping relies on radiation-induced deletions and an appropriate marker platform for detecting those deletions. Improving any of these factors would result in a higher quality map. The induced deletion level is represented by the retention frequency value. An average of 0.5 is the optimum retention frequency for obtaining high quality RH maps (Jones, 2014), although achieving this deletion frequency level is challenging. In animals, RH panel retention frequency is between 0.2-0.3 (Faraut et al., 2009), while in plants it is in the range of 0.7-0.9. The high retention frequency of plant RH panels is because major parts of chromosomes are retained intact when developing *in vivo* populations. The advantage of viable RH panels is that they provide an abundant DNA source. Also, the phenotypic effect of the deleted regions can be analyzed for

detecting underlying genes (Hossain et al., 2004). Furthermore, induced deletions can be retained in the next generation through seed increase.

Applying high radiation dosage improves plant RH maps by decreasing the retention frequency (Kumar et al., 2014). The maximum applied radiation dosage must be determined for each genotype as different genotypes respond differently to radiation treatment. For example, at 35 Krad dosage the germination rate of a synthetic hexaploid wheat and CS was 70% and 21%, respectively (Kumar et al., 2012b; Riera-Lizarazu, 2010). The addition of disomic chromosome 3H made CS more vulnerable to radiation, reducing the germination rate to 0.66% at the same radiation dosage (Figure 2.1). We considered 25 and 15 Krad as the optimum radiation levels for irradiating CS+3H" addition seeds. These radiation dosages; however, did not decrease the retention frequency to less than 0.9 among informative lines. The high retention frequency of 3H-RH lines was compensated by using a large population consisting of 215 informative lines. This population was about three times larger than the RH population of wheat chromosome 3B with the same retention frequency (Kumar et al., 2012a) and resulted in 40-fold higher map resolution.

Due to the uneven deletion distribution, mapping resolution also depends on the number of unique deletions. A high deletion frequency at a particular chromosomal region leads to the low map resolution in that region. For example, 8.79% of lines that lost ≥91.21% of the same region on the long arm did not provide unique deletion patterns for that region, which contributed to the 6.46-fold lower map resolution on the long arm compared to the short arm (Table 2.3). Also, two chromosomal regions, the long arm peritelomeric and the long arm pericentromeric regions, with the highest deletion frequency (Figure 2.3) had the lowest map resolution (Table 2.3). Deletion patterns of contigs anchored along the RH map (Table 2.3)

showed that map resolution depends more on the amount of unique deletions (r=0.62) rather than on retention frequency (r=0.27) (data not shown). Using different radiation dosages might increase the frequency of unique deletions. Based on our data, 15 Krad induced larger deletions that covered similar regions across different lines, whereas 25 Krad generated smaller deletions that were unique to specific lines.

2.4.3.2. Marker platform

Genetic map development is limited by the availability of polymorphic markers. Many markers, especially gene-based markers such as ESTs and candidate genes, are highly conserved among genotypes. For example, in developing a barley genetic map, about half of 7,700 available ESTs were discarded due to a lack of polymorphism (Sato et al., 2009). In contrast, RH mapping is not limited by marker polymorphism. Markers, however, need to be genome-specific to distinguish between the chromosome(s) of interest and the host genome background. Considering the higher genotypic polymorphism among species compared to polymorphism levels of genotypes within single species, designing markers that are polymorphic between different species is more convenient. One approach for obtaining genome-specific markers is designing RJMs (Wanjugi et al., 2009; Mazaheri et al., 2014). Despite the high level of synteny between wheat and barley (Cho et al., 2006), 87.27% of tested RJMs designed from barley sequences did not amplify the wheat background in this study. Also, RJMs are abundant (~400,000 RJMs in the barley genome) and are distributed along chromosomes, covering both coding and non-coding regions (Mazaheri et al., 2014). Abundance and uniform distribution of RJMs make them an ideal marker platform for saturating RH maps through PCR amplification or hybridization-based arrays.

2.5. Conclusion

The 3H-RH map developed in this study is the first RH map of the barley genome. Resolution of the generated map was 2.22 Kb, which is the highest map resolution reported among plant RH maps. This level of resolution enabled anchoring randomly selected contigs, as small as 1.50 Kb in length, onto the generated map. Comparing the skeleton RH map to the analogous genetic map showed a total 9.53-X higher resolution of the RH map, reaching to >262.40-X improved resolution around the centromeric region. We showed that induced deletions were not evenly distributed along chromosome 3H, with the short arm having a 3.37-fold lower deletion frequency compared to the long arm. The uneven distribution of induced deletions led to a maximum of an 8.91-fold deviation from the average map resolution. In general, the middle region of each chromosome arm had the highest resolution (250 bp on the short arm and 1,031 bp on the long arm), and the peritelomeric region had the lowest resolution (819 bp on the short arm and 9,137 bp on the long arm). The generated 3H-RH map, due to the high resolution and the full coverage of all tested chromosomal regions, is an ideal platform for ordering contigs in the assembly of chromosome 3H sequences. The method utilized here can serve as a template for generating RH maps of the remaining barley chromosomes.

2.6. References

Akhunov, E.D., A.W. Goodyear, S. Geng, L.L. Qi, and B. Echalier *et al.* 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. 13:753-763.

Alkhimove, A.G., J.S. Heslop-Harrison, A.I. Shchapova, and A.V. Vershinin. 1999. Rye chromosome variability in wheat-rye addition and substitution lines. Chromosome Res. 7:205-212.

Ashida, T., S. Nasuda, K. Sato, and T.R. Endo. 2007. Dissection of barley chromosome 5H in common wheat. Genes. Genet. Syst. 82: 123-133.

Blake, T., V. Blake, J. Bowman, and H. Abdel-Haleem. 2011. In: S.E. Ullrich, editor, Barley: Production, improvement and uses. Wiley-Blackwell. p. 522-531

Bleuyard, J.Y., M.E. Gallego, and C.I. White. 2006. Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA repair. 5:1-12.

Cho, S., D.F. Garvin, and G.J. Muehlbauer. 2006. Transcriptome analysis and physical mapping of barley genes in wheat-barley chromosome addition lines. Genetics. 172:1277-1285.

Collins, H.M., R.A. Burton, D.L. Topping, M.L. Liao, A. Bacic, and G.B. Fincher. 2010. Variability in fine structures of noncellulosic cell wall polysaccharides from cereal grains: Potential importance in human health and nutrition. Cereal Chem. 87(4):272-282.

Cox, D.R., M. Burmeister, E.R. Price, S. Kim, and R.M. Myers. 1990. Radiation hybrid mapping: a somatic cell genetic method for construction of high resolution maps of mammalian chromosomes. Science. 250:245-250.

De Givry, S., M. Bouchez, P. Chabrier, D. Milan, and T. Schiex. 20005. Cathagene: multipopulation integrated genetic and radiated hybrid mapping. Bioinformatics. 21:1703-1704.

Endo, T.R. 2007. The gametocidal chromosome as a tool for chromosome manipulation in wheat. Chromosome Res. 15:67-75.

Erayman, M., D. Sandhu, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res. 32:3546-3565.

Faraut, T., S. de Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, T. Schiex, B. Servin, A. Vignal, F. Galibert, and M. Yerle. 2009. Contribution of radiation hybrids to genome mapping in domestic animals. Cytogenet. Genome Res. 126:21-33.

Gao, W., Z.J. Chen, J.Z. Yu, D. Raska, R.J. Kohel, J.E. Womack, and D.M. Stelly. 2004. Wide-Cross Whole-Genome Radiation Hybrid Mapping of Cotton (*Gossypium hirsutum* L.). Genetics 167:1317-1329.

Garg, M., H.M.M. Elamein, H. Tanaka, and H. Tsujimoto. 2007. Preferential elimination of chromosome 1D from homoeologous group-1 alien addition lines in hexaploid wheat. Genes Genet. Syst. 82:403-408.

Gaut, B.S. 2002. Evolutionary dynamics of grass genomes. New Phytol. 154:15-28.

Gibbs, R.A., G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, and S. Scherer *et al*. 2004. Genome sequence of the brown Norway rat yields insights into mammalian evolution. Nature. 428:493-521.

Goss, S.J, and H. Harris. 1975. New method for mapping genes in human chromosomes. Nature. 255:680-684.

Guidet, F., P. Rogowsky, C. Taylor, W. Song, and P. Langridge. 1991. Cloning and characterization of a new rye-specific repeated sequence. Genome. 34:81-87.

Hitte, C., J. Madeoy, E.F. Kirkness, C. Priat, T.D. Lorentzen, and F. Senger. 2005. Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. Nature Rev. Genet. 6(8):643-648.

Hossain, K.G., O, Riera-Lizarazu, V. Kalavacharla, M.I. Vales, S.S. Maan, and S.F. Kianian. 2004. Radiation hybrid mapping of the species cytoplasm-specific (scs^{ae}) gene in wheat. Genetics. 168:415-423.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, and D.K. Slonim *et al*. 1995. An STS-based map of the human genome. Science. 270(5244):1945-1954.

Islam, A.K.M.R., K.W. Shepherd, and D.H.B. Sparrow. 1981. Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity. 46:161-174.

Jones, H.B. 2014. Hybrid selection as a method of increasing mapping power for radiation hybrids. Genome Res. 6:761-769.

Joshi, G.P., S. Nasuda, and T.R. Endo. 2011. Dissection and cytological mapping of barley chromosome 2H in the genetic background of common wheat. Genes Genet. Syst. 86:231-248.

Kalavacharla, V., K. Hossain, Y. Gu, O, Riera-Lizarazu, M.I. Vales, S. Bhamidimarri, J.L. Gonzalez-Hernandez, S.S. Maan, and S.F. Kianian. 2006. High-resolution radiation hybrid map of wheat chromosome 1D. Genetics. 173:1089-1099.

Kumar, A., F.M. Bassi, M. Michalak de Jimenez, F. Ghavami, M. Mazaheri, K. Simons, M.J. Iqbal, M. Mergoum, S.F. Kianian, P.M.A. Kianian. Radiation hybrids: A valuable tool for genetic, genomic, and functional analysis of plant genomes. 2014. In: R. Tuberosa, A. Graner, E. Frison, editors, Genomics of Plant Genetic Resources. Springer. Netherlands. p. 285-318.

Kumar, A., F.M. Bassi, E. Paux, O. Al-Azzam, M. Michalak de Jimenez, and A.M. Denton *et al.* 2012a. DNA repair and crossing over favor similar chromosome regions as discovered in radiation hybrid of Triticum. BMC Genomics. 13:339-350.

Kumar, A., K. Simons, M.J. Iqbal, M. Michalak deJimenez, F.M. Bassi, F. Ghavami, O. Al-Azzam, T. Drader, Y. Wang, M.C. Luo, Y.Q. Gu, A. Denton, G.R. Lazo, S.S. Xu, J. Dvorak, P.M.A. Kianian, and S.F. Kianian. 2012b. Physical mapping resources for large

plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genomics. 13:579-593.

Kunzel, G., L. Korzun, and A. Meister. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics. 154(1):397-412.

Kurar, E., J.E. Womack, and B.W. Kirkpatrick. 2003. A radiation hybrid map of bovine chromosome 24 and comparative mapping with human chromosome 18. Anim. Genet. 34:198-204.

Kynast, R.G., R.J. Okagaki, M.W. Galatowitsch, S.R. Granath, M.S. Jacobs, A.O. Stec, H.W. Rines, and R.L. Phillips. 2004. Dissecting the maize genome by using chromosome addition and radiation hybrid lines. PNAS. 101:9921-9926.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, and J. Baldwin *et al*. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Lehmensiek, A., W. Bovill, P. Wenzl, P. Langridge, and R. Apples. 2009. Genetic mapping in the Triticeae. In: C. Feuillet and G.J. Muehlbauer, editors, Genetics and genomics of the Triticeae. Springer, New York. p. 201-235.

Lindblad-Toh, K., C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, and M. Kamal *et al.* 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803-819.

Masoudi-Nejad, A., S. Nasuda, M.T. Bihoreau, R. Waugh, and T.R. Endo. 2005. An alternative to radiation hybrid mapping for large-scale genome analysis in barley. Mol. Gen. Genomics. 274:589-594.

Mayer, K.F.X., R. Waugh, P. Langridge, T.J. Close, R.P. Wise, and A. Graner *et al.* 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature. 491:711-717.

Mayer, K.F.X., M. Martis, P.E. Hedley, H. Simkova, H. Liu, and J.A. Morris, *et al.* 2011. Unlocking the barley genome by chromosomal and comparative genomics. The plant Cell. 23:1249-1263.

Mazaheri M., P.M.A. Kianian, M. Mergoum, G.L. Valentini, R. Seetan, S.M. Pirseyedi, A. Kumar, Y.Q. Gu, N. Stein, M. Kubalakova, J. Dolezel, A.M. Denton, and S.F. Kianian. 2014. Transposable element junctions in marker development and genomic characterization of barley. The Plant genome. 7(1):1-8.

Michalak de Jimenez, M.K., F.M. Bassi, F. Ghavami, K. Simons, R. Dizon, and R.I. Seetan, L.M. Alnemer, A.M. Denton, M. Dogramaci, H. Simkova, J. Dolezel, K. Seth, M.C. Luo, J. Dvorak, Y.Q. Gu, and S.F. Kianian. 2013. A radiation hybrid map of chromosome 1D reveals synteny conservation at a wheat speciation locus. Funct. Integr. Genomics. 13:19-32.

Nasuda, S., Y. Kikkawa, T. Ashida, A.K.M.R. Islam, K. Sato, and T.R. Endo. 2005. Chromosomal assignment and deletion mapping of barley EST markers. Genes Genet syst. 80:357-366.

Nevo, E., Y.B. Fu, T. Pavlicek, S. Khalifa, M. Tavasi, and A. Beiles. 2012. Evolution of wild cereals during 28 years of global warming in Israel. Proc. Natl. Acad. Sci. USA. 109:3412-3415.

Paux, E., P. Sourdille, J. Salse, C. Saintenac, F. Choulet, and P. Leroy. 2008. A Physical map of 1-Gigabase bread wheat chromosome 3B. Science. 322:101-104.

Pitel, F., B. Abasht, M. Morisson, R.P.M.A. Crooijmans, F. Vignoles, S. Leroux, K. Feve, S. Bardes, D. Milan, S. Lagarrigue, M.A.M. Groenen, M. douaire, and A. Vignal. 2004. A high- resolution radiation hybrid map of chicken chromosome 5 and comparison with human chromosomes. BMC Genomics. 5:66-74.

Riera-Lizarazu, O., M.I. Vales, E.V. Ananiev, H.W. Rines, and R.L. Philips. 2000. Production and characterization of maize chromosome 9 radiation hybrids derived from an oat-maize addition line. Genetics. 156:327-339.

Riera-Lizarazu, O., J.M. Leonard, V.K. Tiwari, and S.F. Kianian. 2010. A method to produce radiation hybrids for the D-genome chromosomes of wheat (*Triticum aestivum* L.). Cytogenet. Genome Res. 129:234-240.

Rudi, H., A.K. Uhlen, O.M. Harstad, and L. Munck. 2006. Genetic variability in cereal carbohydrate compositions and potentials for improving nutritional value. Anim. Feed Sci. Technol. 130:55-65.

Sakai, K., S. Nasuda, K. Sato, and T.R. Endo. 2009. Dissection of barley chromosome 3H in common wheat and a comparison of 3H physical and genetic maps. Genes Genet. Syst. 84:25-34.

Sakata, M., S. Nasuda, and T.R. Endo., 2010. Dissection of barley chromosome 4H in common wheat by the gametocidal system and cytological mapping of chromosome 4H with EST markers. Genes. Genet. Syst. 85: 19-29.

Sato, K., Y. Motoi, N. Yamaji, and H. Yoshida. 2011. 454 sequencing of pooled BAC clones on chromosome 3H of barley. BMC Genomics. 12:246-252.

Sato, K., N. Nankaku, and K. Takeda. 2009. A high-density transcript linkage map of barley derived from a single population. Heridity. 103:110-117.

Serizwa, N., S. Nasuda, F. Shi, and T.R. Endo. 2001. Deletion-based physical mapping of barley chromosome 7H. Theor Appl. Genet. 103: 827-834.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, and K. Rice *et al.* 1996. A gene map of human genome. Science. 274 (5287):540-546.

Stephens, J.L., S.E. Brown, N.L.V. Lapitan, and D.L. Knudson. 2004. Physical mapping of barley genes using an ultrasensitive fluorescence in situ hybridization technique. Genome. 47:179-189.

Stewart, E.A., K.B. McKusick, A. Aggarwal, E. Bajorek, S. Brady, and A. Chu *et al.* 1997. An STS-based radiation hybrid map of the human genome. Genome Res.7:422-433.

Taketa, S., J. Kato, and K. Takeda. 1995. High crossability of wild barley (*Hordeum spontaneum* C. Koch) with bread wheat and the differential elimination of barley chromosomes in the hybrids. Theor. Appl. Genet. 91(8):1203-1209.

Wanjugi, H., D. Coleman-Derr, N. Huo, S.F. Kianian, M.C. Luo, J. Wu, O. Anderson, and Y.Q. Gu. 2009. Rapid development of PCR-based genome specific repetitive DNA junction markers in wheat. Genome. 52(6):576-587.

Wardrop, J., J. Snape, W. Powell, and G.C. Machray. 2002. Constructing plant radiation hybrid panels. Plant J. 31(2):223-228.

Wenzl, P., H. Li, J. Carling, M. Zhou, H. Raman, and E. Paul. 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics. 7: 206-228.

Womack, J.E., J.S. Johnston, E.K. Owens, C.E. Rexroad III, J. Schlapfer, and Y.P. Yang. 1997. A whole-genome radiation hybrid panel for bovine gene mapping. Mamm. Genome. 8:854-856.

You F.M, H. Wanjugi, N. Huo, G.R. Lazo, M.C. Luo, O.D. Anderson, J. Dvorak, and Y.Q. Gu. 2010. RJPrimers: unique transposable element insertion junction discovery and PCR primer design for marker development. Nucleic Acids Res. 38:313-320.

3. TRANSPOSABLE ELEMENT JUNCTIONS IN MARKER DEVELOPMENT AND GENOMIC CHARACTERIZATION OF BARLEY

Barley (*Hordeum vulgare* L.) is a model plant in genomic studies of Triticeae species. However, barley large genome size and high repetitive sequence content complicate the whole-genome sequencing .The majority of the barley genome is composed of transposable elements (TEs). In this study, TE repeat junctions (RJs) were utilized to develop a large-scale molecular marker platform, as a prerequisite to genome assembly. A total of 10.22 Gb of barley non-assembled 454 sequencing data were screened with RJPrimers pipeline. In total, 981,561 TE junctions were identified. From detected RJs, 400,538 PCR-based RJ markers (RJMs) were designed across the genome, with an average of 39 markers per Mb. The utility of designed markers was tested using a random subset of RJMs. Over 94% of the markers successfully amplified amplicons, among which ~90% were genome specific. In addition to marker design, identified RJs were utilized to detect 1,190,885 TEs across the genome. In gene-poor regions of the genome *Gypsy* elements comprised the majority of TEs (~65%), while in gene-rich regions *Gypsy*, *Copia*, and *Mariner* were the main transposons, each representing an average ~23% of total TEs. The numerous RJ primer pairs developed in this study will be valuable resource for barley genomic studies including genomic selection, fine mapping, and genome assembly. In addition, the results of this study show that characterizing RJs provides insight into TE composition of species without a sequenced genome but for which short-read sequence data is available.

3.1. Introduction

Barley, (*Hordeum vulgare* L.*,* 2n=2x=14) ranking fifth in the world food production (FAOSTAT, 2007; http://faostat.fao.org/), is an important cereal crop cultivated worldwide. Barley is the major raw material for the brewing industry and for animal feed; also, it offers

considerable health benefits for the human diet (Baik and Ullrich, 2008). As a diploid and self-pollinated crop exhibiting high collinearity with other Triticeae species, barley is considered a model plant for genetic studies of related species with complex genomes such as wheat and rye. For these reasons, efforts are underway to fully sequence the barley genome (http://barleygenome.org/). A complete sequence of the barley genome will provide valuable information about the genetic structure, especially as related to important agronomic traits. This information can be exploited to improve barley and other related Triticeae species.

The barley genome is large (5.1 Gb) and is comprised of ~84% repetitive sequences (The international barley genome sequencing consortium, 2012). The large amount of repetitive elements is a major challenge to genome assembly and results in large gaps between sequenced fragments. Recently, deep shotgun sequencing of the barley genome resulted in the assembly of contigs totaling 1.9 Gb in length. The remainder of the genome failed to assemble due to the high percentage of repetitive sequences (The international barley genome sequencing consortium, 2012). Development of a high-resolution marker scaffold will facilitate genome assembly by anchoring contigs and repetitive sequence reads to chromosomes. In this effort TEs, constituting more than 98% of the repetitive sequences in the barley genome (The international barley genome sequencing consortium, 2012), provide an excellent opportunity for large-scale marker development due to their wide distribution across the genome (Kalendar et. al., 2011; Kumar et al., 2012). TEs are repetitive sequences that can change their location within the genome and are categorized into two main classes depending upon the mode of action: Class I, or retrotransposons, move in a ‘‘copy-and-paste” mechanism by amplifying through intermediate RNA and inserting into new regions; Class II, or DNA transposons, move within the genome by excising from their original location and inserting into a new region in a ‘‘cut-and-paste’’

mechanism. Each TE class is further classified into subclasses, orders, and superfamilies (Wicker et al., 2007). It is established that transposon activity is a main factor in inducing genomic diversity and evolution of species (Feschotte et al., 2002). Therefore, in contrast to coding sequences, which are conserved among related species, transposon arrangements are variable, even among close lineages.

TEs abundance and polymorphism have been exploited to develop several classes of molecular markers, most of which utilize TE insertion sites (for a review see Kalendar et al., 2011). TE insertions generate unique junctions comprised of the boundaries between TEs and sequences of loci in which they land (Bennetzen, 2000). Recently, You et al. (2010) developed RJPrimer software for the automated design of five classes of TE junction-based markers including: repeat junction markers (RJM), repeat junction-junction markers (RJJM), insertion-site-based polymorphism (ISBP), inter-retrotransposon amplified polymorphism (IRAP), and retrotransposon-based insertion polymorphism (RBIP). Among these markers, ISBP and RJM are useful for high-throughput genotyping because they generate the highest number of primer pairs (You et al., 2010). However, RJMs provide an extra advantage over ISBPs in functional genomic studies because ISBP markers are solely designed based on TE sequences, whereas RJMs could span both TE and genic sequences and are able to capture genes. RJM design involves identifying one primer that spans the unique TE junction sequence and another primer from any type of sequence (Devos et al., 2005). The uniqueness of RJMs has been utilized to map the D genome of hexaploid wheat by designing primers from its wild donor (*Aegilops tauschii,* DD) (Wanjugi et al., 2009).

In addition to developing molecular markers, RJ analysis reveals transposon target regions that could unravel the mechanisms underlying dynamic evolutionary genome changes. It

was previously shown that TE insertion is not a random process and some transposons are preferentially inserted into specific regions (Li et al., 2004; Paux et al., 2006). However, characterization of TE insertion has been mainly limited to few bacterial artificial chromosomes (BACs) that were manually annotated (Li et al., 2004; Paux et al., 2006; Bartos et al., 2008; Wicker et al., 2009). In recent years, advances in next-generation sequencing have drastically reduced sequencing costs, resulting in the availability of vast amounts of sequencing data (Metzker, 2010). These resources confer excellent opportunities to study genomes of species that are not completely sequenced.

In this study, we used barley survey sequencing data to detect RJs across the barley genome. Based on the detected RJs: 1) a large number of RJMs were designed, and their efficiency in PCR amplification was tested, 2) the insertion pattern of TEs was characterized, and 3) the TE composition from survey sequencing data was compared with that from gene-rich regions. The results of this study illustrate the utility of RJs toward analyzing highly repetitive and complex genomes and provide an overview of TE arrangements in the barley genome.

3.2. Materials and Methods

3.2.1. Genome sequences

Shotgun sequencing data of barley chromosome 1H and 12 chromosomes arms (2HS-7HL), generated from flow-sorted chromosomes, was downloaded from the NCBI SRA database (http://www.ncbi.nlm.nih.gov/sra?term=hordeum%20vulgare%20[orgn]%20chromosome). The accession used for sequencing was barley cultivar Betzes. In total 10.22 Gb sequencing data with 1.04-2.00X coverage was downloaded to detect RJs and to design RJMs across the genome. Available sequences of EST-based contigs from chromosome 3H were used to extract TE composition from gene-rich regions. These sequences were also used to design 96 RJMs and test

their efficiency for PCR amplification. The selected contigs originated from BAC clone sequences containing ESTs with known positions on a saturated genetic map developed by Sato et al. (Sato et al., 2011). From each map position, the largest contigs were selected. In total, 172 contigs with an average size of 19,490 bp were used. Contig sequences were downloaded from the Gbrowse system at Barley Contig Viewer (http://150.46.168.145/gbrowse/).

3.2.2. Repeat junction analysis and developing repeat junction markers

RJs were detected by RJPrimers pipeline v1.0 (http://probes.pw.usda.gov/RJPrimers/). The Triticeae repeat (TREP) sequence database was used for BLAST analysis and RJ detection. E-value cutoff was set to 1e-50 for the minimum top hit and 1e-5 for the minimum of all hits to reduce the detection of false positive RJs. Program default settings were used to design RJ primers from the detected RJs. To test for TE insertion specificity, chi square test at $p<0.001$ was performed. The expected frequency of each TE junction was calculated by multiplying the total number of junctions for the row and column of the desired cell (TE) and then dividing by the total number of observed junctions, as described by Boslaugh and Watters (2008). The formula can be represented as follows: Expected frequency of cell (ij) $= \frac{i^{th}\text{row totals} \times j^{th}\text{column totals}}{\text{total number of observed junctions}}$

3.2.3. Detection of transposable elements

A script was written to detect TEs based on junction incidence. The developed script extracts TE copy numbers from detected RJs and their associated TEs, provided by RJPrimers BLAST search. The script is 99% accurate in detecting TEs when combined with RJPrimer output. TE identification was not possible in less than 1% of analyses due to lack of TE and RJ positional information. Identified TEs were categorized based on the classification of RJPrimers default parameters.

3.2.4. DNA materials

PCR amplification was conducted on seven barley lines and one wheat cultivar (*Triticum aestivum)* cultivar "Chinese Spring". The selected barley lines included five barley cultivars (*Hordeum vulgare*): "Betzes", "Golden Promise", "Bowman", "Morex", "Haruna Mugi", and two wild barley (*H. bulbosum*) accessions PI 106880 and PI 206565, from Turkmenistan and Greece, respectively. Genomic DNA was extracted from the lyophilized tissue of young leaves, as described by Guidet et al. (Guidet et al., 1991). To estimate the chromosome specificity of RJMs, entire chromosome 1H and 12 chromosome arms (2HS to 7HL) were isolated by flow cytometric sorting from cv. "Betzes" and from wheat-barley ditelosome addition lines, respectively, as described by Suchánková et al. (2006). DNA of flow-sorted chromosomes was purified and amplified by Phi29 polymerase, as described by Šimková et al. (2008).

3.2.5. PCR amplification

Ninty-six RJMs distributed along chromosome 3H were randomly selected from gene-based contigs to amplify genomic or chromosomal DNA. PCR amplification was conducted in 20 µl volumes consisting of 45 ng of genomic DNA or 4 ng of chromosomal DNA, 2.5 mM of $MgCl_2$, 1X PCR buffer, 0.25 mM of dNTPs, 0.37 µM of each primer, and 1.0 unit of *Taq* DNA polymerase. Touchdown PCR program was used as follows: 94°C for 4 min., 5 cycles of 94°C for 30 sec., 65°C for 30 sec., 72°C for 1 min., with the temperature decreased 1°C per cycle, followed by 35 cycles of 94°C for 30 sec., 60°C for 30 sec., 72°C for 1 min., and 72°C for 7 min. PCR products were visualized on agarose (1.5% Agarose S, SeaKem ®LE Agarose) gel electrophoresis (110 V, 45 min.) and scored for presence/absence. All the markers were multiplexed with a chloroplast-specific marker to provide a positive control against PCR miss-

amplification. Polymorphic amplicons were applied to cluster barley lines by the DARwin 5.0.158 NJ unweighted clustering method.

3.2.6. Detection of SNPs and indels in RJMs

PCR products of six RJMs were purified using the QIAquick®Gel Extraction kit. Purified fragments were pair-end sequenced by GENWIZ Sanger DNA Sequencing Services. Sequence data were manually inspected for SNP and indel presence using ChromasPro v1.5 software. Sequences with low-quality reads were omitted, and only the most reliable polymorphic nucleotides were determined as SNPs/single nucleotide indels.

3.3. Results

3.3.1. Development of large scale repeat junction markers

3.3.1.1. Designing RJ primers across the barley genome

A total of 10.22 Gb unassembled 454 survey sequencing data were parsed for the presence of RJs using RJPrimers pipeline (Wanjugi et al., 2009). In total, 981,561 RJs with an average of 96 RJs per Mb were identified. It was assumed that RJs near the ends of 454 short reads (with average length of 482 bp) were undetected. To validate this assumption, we estimated the frequency of RJs based on 172 available chromosome 3H BAC contigs, with an average length of 19,490 bp. The frequency of RJs detected from contig sequences was 512/Mb, which is five times higher than the frequency of RJs detected in the 454 short-read sequences. Thus, it is anticipated that the actual number of RJs in the barley genome is about five times higher than what was estimated using shotgun survey sequences. The number of detected RJs was similar among all chromosomes except for chromosome 1H, where the number of detected RJs was almost half that of the other chromosomes. The small number of RJs identified on

chromosome 1H was likely due to the short length of 1H sequence reads, which were half the size of other chromosome reads (Table 3.1).

About 40.8% of identified junctions were developed into RJ primer pairs using RJPrimers software. The remaining junctions failed RJPrimer minimum threshold parameters designed to increase the likelihood of PCR amplification. Overall, 400,536 RJMs were developed from the 454 survey sequencing data, with an average of 39 markers per Mb (Table 3.1).

Table 3.1. Number of detected RJs and RJ primer pairs per chromosome

Source	Length of sequence reads (Mb)	Number of detected RJs	Number of RJs per Mb	Number of developed RJ primer pairs	Number of RJ primer pairs per Mb
1H	790.63	44968	57	15970	20
2H	1452.18	156049	107	64502	44
3H	1823.77	158410	87	68451	38
4H	1564.32	143192	92	58859	38
5H	1708.73	174219	102	71011	42
6H	1606.57	158904	99	64519	40
7H	1275.54	145819	114	61797	48
Whole genome	10221.74	981561	96	400536	39

3.3.1.2. Validation of RJMs in PCR amplification

To test RJM utility, 35 RJ primer pairs distributed throughout chromosome 3H of barley cultivar "Haruna Mugi" were designed. Selected primer pairs were used in PCR reactions on seven barley lines and one wheat cultivar ("Chinese Spring"), as negative control. An addition set of 61 RJMs were tested on a radiation hybrid (RH) population of Betzes chromsome 3H and Chinese Spring (unpublished data). In total from 96 tested markers, 94.64% successfully amplified one, or in few cases two, amplicons, among which 90.31% of the amplicons were barley-specific. Among 35 markers tested on barley accessions, 63.15% of the amplicons were polymorphic as presence/absence. This level of polymorphism separated all seven barley lines (Figure A.1). To test RJM chromosome specificity, the same 35 primer sets were used to amplify flow-sorted chromosome arms of barley (cultivar "Betzes"). All primer sets amplified DNA

fragments from chromosome 3H, except one primer pair that amplified a fragment from chromosome 5H, instead. This result showed that RJMs designed from a specific chromosome of one cultivar can be applied to the same chromosome of a different cultivar. Assuming no contamination in chromosome sorting, 22.85% of the amplicons were specific for chromosome 3H, 68.58% were common to chromosome 3H and another chromosome(s), and 8.57% produced amplicons from all seven chromosomes.

To estimate single nucleotide polymorphism (SNP) and insertion/deletion (indel) rates within the RJM amplicons, six RJMs were selected: three RJMs with polymorphic and three RJMs with monomorphic amplicons among the seven barley lines. Amplicons from each line were sequenced for each selected marker. In total, 10,426 bp of sequence data were collected, and an average of one SNP/single nucleotide indel per 652 bp was detected among which 25.01% were single nucleotide indels and 74.99% were SNPs. Transition and transversions comprised 55.55% and 44.45% of SNPs, respectively. The larger portion of transition to transversion frequency was expected due to high levels of TE methylation (Clark et al., 2005; Rabinowicz et al., 2005). BLAST analysis of amplicon sequences against the TIGR EST barley database indicated that three of the six non-TE sequences contained genic sequences. It should be considered that in this study RJMs were characterized by using a small number of markers. To have a better estimation of polymorphism, chromosome specificity, and SNP rate large numbers of RJMs need to be tested.

3.3.2. Characterizing transposable elements in the barley genome using repeat junctions

3.3.2.1. Insertion pattern of transposons in the genome

Detection of RJs allowed us to analyze the insertion pattern of TEs in the barley genome. Identified junctions were classified into five RJ classes based on the inserted elements (DNA

transposons or retrotransposons) and target regions (DNA transposons, retrotransposons, or non-TE elements). In this study, non-TE elements refer to target sequences that are mainly composed of genes or intergenic regions (You et al., 2010). Target and inserted elements were not differentiated, and RJs were classified solely on the two elements involved in the junctions, regardless of the order. For example, both "DNA transposon-retrotransposon" and "retrotransposon-DNA transposon" junctions were categorized into one class. Frequency of the five RJ classes and the most abundant junctions of TE superfamilies are shown in Figure 3.1. Insertion of retrotransposons, primarily *Gypsy* and *Copia*, into non-TE sequences comprised the majority of junctions (65.43%). "Retrotransposons-retrotransposons" comprised the second most common junction (19.52%), mainly containing the association of *Gypsy* with *Gypsy* and *Gypsy* with *Copia*. The third most common junction (10.15%) resulted from the insertion of DNA transposons, mostly *CACTA* and in a lower amount *Mariner*, into non-TE regions. The least common junctions were "DNA transposon-retrotransposon" and "DNA transposons-DNA transposons", representing only 3.76% and 1.15% of RJs, respectively. However, it should be considered that TE arrangements observed in current cultivars have been undergone evolutionary changes such as genomic rearrangement and selection. Therefore, TE arrangements detected in this study do not necessarily represent ancestral TE insertion patterns in the barley genome.

Preferential association of nested TEs was explored by comparing the frequency of each TE-TE junction occurrence relative to an expected frequency of occurrence. Expected frequency of occurrence refers to the junction frequency if TE insertion occurs randomly, based on the relative abundance of target and inserted elements. Our analysis showed that TE insertion is not random, and TE superfamilies have preferential associations with one another (Table A.1). Overall, there were preferential associations within class I and within class II elements, with the

exception of LINE (class I) that had more association with *CACTA* and *Mariner* elements (class II) and less association with *Gypsy* (class I).

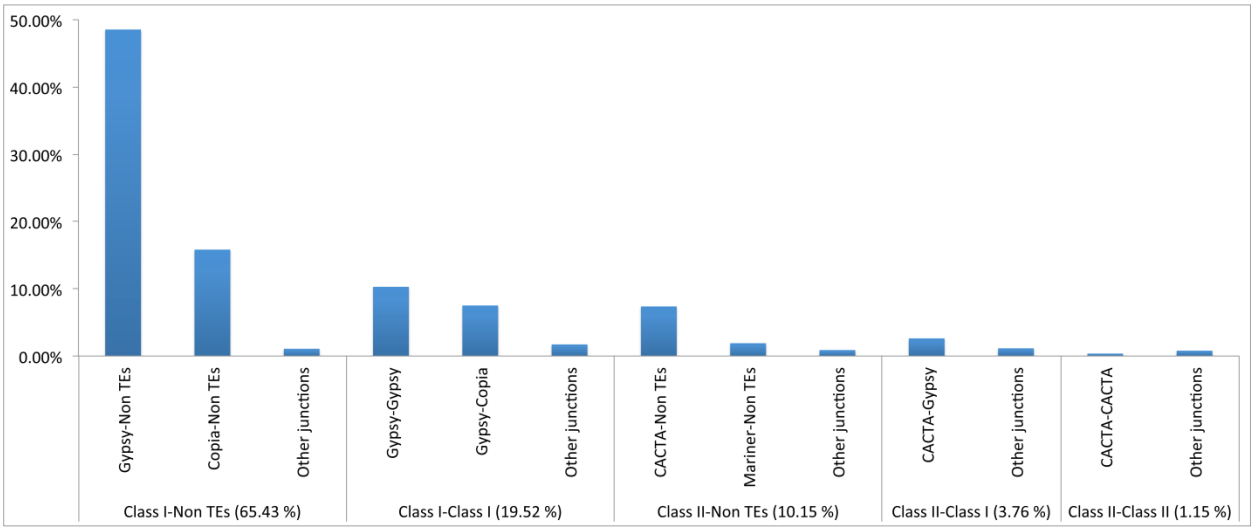

Figure 3.1. Frequency of five RJ categories (Class I-Non TEs, Class I-Class I, Class II-Non TEs, Class II-Class I, and Class II-Class II) and the most frequent superfamily junctions in the barley genome

3.3.2.2. Estimating copy number of transposons by their insertion sites

We developed an algorithm to identify TE copy number based on the unique junction sequences detected by RJPrimers. This software is specifically designed to detect RJs, and it does not provide accurate information on TE copy number. For example, if two RJs are detected at both ends of one TE in a short sequence, the TE will be reported twice in the RJPrimers output file. Consequently, extracting TEs from the output file would overestimate the TE copy number. To overcome this drawback we developed a script that counts the correct copy number of TEs based on the information provided by the RJPrimers BLAST search. For example, considering the segment "non TE-*CACTA*-non TE", two RJs (*i.e.*, non TE–*CACTA*, *CACTA*-non TE) and two *CACTA* elements (*i.e.*, same element, counted once at each of two junctions) are detected by RJPrimers. The program also provides information on the position of the two RJs on the sequence read and length of detected *CACTA*. Based on the first RJ position and *CACTA* length,

our script estimates the position of the second RJ. If the second RJ position, estimated by the script, matches the second RJ position provided by RJPrimers, then the two *CACTA* elements are considered as one. Consequently, the script records two RJs but only one *CACTA*. This process is done for each read separately. The script is available upon request. Utilizing this method on the barley survey sequencing data, 1,190,885 TEs were detected, with an average of 117 per Mb. TE compositions were similar among all seven chromosomes (Table A.2).

3.3.2.3. Transposon composition in gene-rich regions

To further investigate the distribution of TEs in the genic regions, we narrowed the focus of analysis from 10.22 Gb of the survey sequencing data to 3.35 Mb sequences of chromosome 3H contigs anchored to EST markers (Sato et al., 2011). At the time of this study, the selected contigs were the longest sequences anchored to a genetic map. The majority of markers on genetic maps were previously shown to represent a small portion of the Ttiticeae genome located in recombination hot spot regions (Künzel et al., 2000, Erayman et al., 2004). Also, it has been shown that genes are mainly clustered in small gene islands within the Triticeae genome (Wicker et al., 2001; SanMiguel et al., 2002; Gu et al., 2004, Lehmensiek et al., 2009). Therefore, we can assume that the selected EST-anchored contigs mainly represent small genomic regions composed of recombination hot spots within gene-clustered islands. Based on this assumption, we considered the selected contigs as representative of gene-rich regions compared to the random 454 survey sequencing data that represent the majority of the genome, which is relatively poorer in genic sequence content. Our data showed substantial differences in TE composition between gene-rich regions of chromosome 3H and the survey sequencing data of the same chromosome (Table 3.2).

Table 3.2. Comparison of TE copy numbers in EST-based contigs and 454 random reads of chromosome 3

Class	Order	Superfamilies	EBC[1]	RR[2]
DNA transposons (EBC=44.01%) (RR=12.96%)	TIR (EBC=42.96%) (RR=12.92%)	*CACTA*	12.15%	9.74%
		Mariner	22.87%	1.91%
		Mutator	3.16%	0.46%
		hAT	0.38%	0.01%
		Harbinger	4.40%	0.45%
		Unclassified	0.00%	0.35%
	Helitron	*Helitron*	0.67%	0.02%
	Unclassified	-	0.38%	0.02%
Retrotransposons (EBC=55.97%) (RR=87.04%)	LTR (EBC=48.89%) (RR=86.71%)	*Gypsy*	24.78%	64.71%
		Copia	21.72%	21.00%
		Unclassified	2.39%	1.00%
	LINE	Unclassified	6.89%	0.32%
	SINE	Unclassified	0.19%	0.01%

[1]EBC=EST based contigs, [2]RR= 454 random reads

In gene-rich regions, DNA transposon content increased 3.39 times from 12.97% in the survey sequencing data to 42.97%. This increase was mainly due to the 12-fold increase of *Mariner* superfamily in gene-rich regions. Conversely, the proportion of retrotransposon copy number decreased drastically from 86.03% in gene-poor to 55.98% in gene-rich regions, mainly due to a 2.6-fold reduction of *Gypsy* superfamily in gene-rich regions. In the survey sequencing data, TEs were mainly composed of three elements: *Gypsy*, *Copia*, and *CACTA*, covering 65.42%, 21.64%, and 9.43% of transposons, respectively. In contrast, TE composition of gene-rich regions comprised four dominant elements: *Gypsy*, *Copia*, *Mariner*, and *CACTA*, where *Gypsy*, *Copia*, and *Mariner* each represented an average 23.12% of TEs, and *CACTA* comprised 12.15% of TEs. Moreover, in gene-rich regions, retrotransposons-LINE, DNA transposon-

Harbinger, and DNA transposon-*Mutator* content increased 21.53-, 9.77-, and 6.86-fold, respectively.

The majority of TE superfamilies were distributed evenly along chromosome 3H, with the exception of two superfamilies. The first exception was retrotransposon-*Copia* that was significantly more concentrated in the pericentromeic region and less concentrated in the short arm sub-telomeric region. The other exception was DNA transposon-*CACTA* that was significantly more concentrated on the long arm telomeric region. (Figure A.2 and Table A.3). However, the distribution of TEs identified in this study resulted from the analysis of sequences anchored to a genetic map. To identify the precise chromosomal regions occupied by specific TEs, transposon distribution needs verification on physical maps that are anchored along chromosomes.

In gene-rich regions of chromosome 3H most of the detected RJs (97.38%) were the result of TE insertions into non-TE sequences (Table A.4). This result was expected because the frequency of genic and intergenic sequences, or non-TEs, is high in the gene-rich regions. Therefore, it is reasonable to detect the majority of TEs in the vicinity of non-TE sequences. Insertion of TEs into other TEs (nested TEs) in gene-rich regions composed only 2.62% of detected junctions, mostly located on the pericentromeric and long arm proximal regions (Figure A.3).

3.4. Discussion

3.4.1. RJMs in genome analysis

Analysis of Triticeae genomes has been a major challenge due to their large size and high repetitive sequence content. In this study, we developed several hundred thousand RJMs distributed across the barley genome. Designed RJMs are potentially a valuable genetic tool for

numerous genomic studies of the barley genome, including complete genome sequencing, map-based cloning, marker-assisted selection (MAS), genetic diversity analysis, evolutionary studies, and RH mapping.

RJMs could be indispensably useful for anchoring available barley sequence reads to chromosomes and developing a complete genome sequence (unpublished data). The recently published barley draft sequence is composed of physical maps that successfully cover most of the genome; however, within physical maps considerable amount of sequence reads failed to assemble due to the high repetitive element content (The international barley genome sequencing consortium, 2012). In our study, an average 96 RJs/Mb were detected from unassembled 454 survey sequencing data. It was estimated that density of detected RJs can be increased to more than 480 RJs per Mb by using longer sequences of assembled reads. We were able to convert half of the identified RJs to PCR-based primer pairs; however, potentially all detected RJs could be utilized as probes in hybridization-based high throughput genotyping systems. The high density, wide distribution, and distinctiveness of RJMs could greatly improve the assembly of sequenced fragments and bridge the genomic gaps.

Additionally, RJMs are valuable for positional cloning, MAS, and saturating available genetic maps. We observed that TEs have the tendency to insert into specific regions. For example, 92.29% of *Mariner* elements were found in the gene-rich regions (Table 3.2), making it likely that *Mariner*-based RJMs map adjacent to genes. In contrast, *Gypsy*-based RJMs could be suitable for whole genome analysis, as *Gypsy* superfamily is the most abundant transposon in the barley genome. Also, we showed that *Copia* and *CACTA* occupy different chromosomal regions (Figure A.2 and Table A.3). Therefore, distinct classes of RJMs can be utilized according to the purpose of a given study.

RJMs can also elucidate genetic diversity and pedigree relationships. It is well established that environmental stress triggers TE activity (Hua-Van et al., 2011). Although TE-induced mutations can be lethal, in certain situations they improve genome adaptation to environmental conditions. Consequent evolutionary forces such as genetic drift can lead to the generation of new genotypes within species (Hua-Van et al., 2011). Therefore, it is reasonable to expect high levels of RJM polymorphism among even closely related genotypes compared to other molecular markers. RJMs used in this study produced 63.15% polymorphic amplicons, which facilitated differentiation of all seven tested lines. However, the rate of polymorphism was different among different RJMs. We found that "*Stowaway*-non-TE" markers had the highest polymorphism (87.50%), in contrast; "*CACTA-CACTA*" markers were monomorphic among all tested lines. Among RJMs, retrotransposon-based markers are highly informative toward recognizing ancestral lines in phylogenetic trees. Retrotransposons amplify within the genome while they retain the original copies/locations. As a result, the absence of a retrotransposon could point toward the ancestral state (Kalendar et al., 2011). This situation was observed in our data, where 76.92% of retrotransposon-based RJMs were absent in the two wild relatives of barley and present in each of the five cultivars screened.

RJMs could also be valuable for physical map development through RH mapping (Kumar et al., 2012). The main obstacle in RH mapping is identifying genome-specific markers. For example, in developing a RH map of barley, mapping radiation-induced breaks of barley chromosomes in a wheat background is a conventional practice. Considering the high synteny between wheat and barley genomes, finding markers specific for barley that do not amplify wheat genomic sequence is a major challenge. Testing barley EST markers showed that only 14.21% of the markers were polymorphic between hexaploid wheat (cultivar "Chinese Spring")

and barley (cultivar "Betzes") (Nasuda et al., 2005), while amplification of RJMs on the same lines resulted in a 92.10% polymorphism rate. In addition, the deletion frequency of RJMs in the RH panel was reported about three times higher than the deletion rate of SSR and gene-based markers (Kumat et al., 2012). We expect that RJMs will increase RH mapping efficiency in the near future.

3.4.2. RJ analysis could reveal the structure of the genome

The majority of detected RJs (75.57%) from the survey sequencing data resulted from transposon insertions into non-TE sequences (Figure 3.1). This result was expected since it is known that the majority of Triticeae genome is composed of dispersed TEs (Wicker et al., 2011, SanMiguel et al., 2002, Gu et al., 2004). A more detailed analysis of RJs revealed that the majority of the barley genome is composed of strings of retrotransposons (*Gypsy*, and to a lesser extent, *Copia*) and coding regions that are alternately repeated throughout the genome. Nested transposons comprised 24.42% of the RJs, mainly resulting from the insertion of *Gypsy* elements into other *Gypsy* elements, or from the association of *Gypsy* and *Copia* (Figure 3.1). Also, whole-genome analysis of RJs revealed that insertion of TEs into other TEs is not random, and most TE superfamilies have preferential target transposons (Table A.1). Transposons with low copy number (*i.e.*, DNA transposons and non-LTR retrotransposons) preferentially insert into low copy number TE superfamilies, and high copy number transposons, *Gypsy* and *Copia,* preferentially target themselves (*i.e.*, *Gypsy* to *Gypsy*, and *Copia* to *Copia*) (Table A.1).

3.4.3. Unequal distribution of TEs in the genome

We developed a script to identify TE copy numbers based on their junctions. The advantage of discovering TEs by their unique insertion sites, compared with available homolog-based programs, is the ability to detect TEs in unassembled short sequence reads. This approach

can also facilitate detection of new TEs with structures that have changed over time. Applying our script, coupled with RJ Primers output data, enables TE detection in unassembled short-reads. Using this method, we were able to detect 1,190,885 TEs in 10.22 Gb of barley 454 sequencing data. However, it should be considered that the number of TEs obtained in this study is an estimate based on the available unassembled sequences. This number would increase by using long sequence reads or decrease by discarding the redundant TEs. The proportion of TE superfamilies obtained in this study was comparable with the TE proportions observed in assembled contigs (The international barley genome sequencing consortium, 2012) illustrating the efficiency of detecting TEs by their insertion sites in unassembled short reads.

Using this technique, we determined that TE distribution is similar among all chromosomes (Table A.2) but considerably different among gene-rich and gene-poor regions. In the gene-rich regions *Gypsy* copy number dropped by a factor of 2.61. The sudden drop of *Gypsy* is likely a consequence of selection pressure against the possibly deleterious insertion effect on gene function. By contrast, in gene-rich regions DNA transposons and non-LTR retrotransposons (LINE and SINE) increased by factors of 3.39 and 7.01, respectively. This result was expected since most DNA transposons and low-copy number retrotransposons are clustered in gene-rich regions in maize and wheat (Bureau et al., 1994; Li et al., 2004; Bruggmann et al., 2006; Paux et al., 2006). The most drastic increase of DNA transposons was observed as a 12-fold increase of *Mariner* (Table 3.2). The high concentration of this superfamily in critical gene-rich regions might indicate the important role of *Mariner* toward improving gene functions. *Copia* was the only TE for which copy number did not change between gene-rich and gene-poor regions. This superfamily was more concentrated in centromeric and pericentromeric regions, which may indicate its role in centromerization or heterochromatinization.

Previously, it was suggested that unequal distribution of TEs could be explained by the gradual increase of recombination frequency from centromere to telomeres (Li et al., 2004). However, we found a different distribution of *Copia* and *CACTA* between short and long chromosome arms (Figure A.2 and Table A.3), which indicates that factors other than recombination are involved in the distribution of TEs. Differences in preferential TE target sites and selection pressure on the inserted elements could be two factors in variable TE distribution on chromosomes.

3.5. Conclusion

In this study, we utilized unique RJ sequences to develop 400,536 PCR-based RJ primer pairs across the barley genome. The abundance of developed markers makes them useful for the barley genome assembly, as well as MAS and genomic selection in breeding programs. Also, due to the high genome specificity (90.13%), RJMs are ideal for RH mapping of the barley genome, currently underway. In addition to designing markers, we utilized RJ sequences to detect 1,190,885 TEs and their preferential insertion sites in 10.22 Gb of unassembled 454 survey sequencing data. The results of this study provide a general picture of TE arrangements in the barley genome and represent the first whole-genome analysis of TE insertion pattern in Triticeae. We showed that the majority of the barley genome was composed of alternate repeats of *Gypsy* and genic sequences. However, in gene-rich genomic regions *Gypsy* frequency decreased, while the frequency of *Mariner* and *Copia* increased, where each of the three elements comprised ~23% of transposons. Also, the majority of TEs in the gene-rich regions were distributed evenly along chromosome 3H, except for *Copia* and *CACTA*, which had variable distribution. We also showed that the majority of TE superfamilies were not inserted into random transposons but tended to insert into specific TEs. The results of this study indicate

the efficiency of RJ sequences in designing high throughput markers and characterizing TEs in species for which the complete genome sequence is not available.

3.6. References

Baik, B.K., and S.E. Ullrich. 2008. Barley for food: characteristics, improvement, and renewed interest. J Cereal Sci. 48**:**233-242.

Bartos, J., E. Paux, R. Kofler, M. Havrankova, D. Kopecky, P. Suchankova, J. Safar, H. Simkova, C.D. Town, T. Lelley, C. Feuillet, and J. Dolezel. 2008. A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biol. 8:95-106.

Bennetzen, J.L. 2000. Transposable element contributions to plant gene and genome evolution. Plant. Mol. Biol. 42:251-269.

Boslaugh, S., and P.A. Watters. 2008. Categorical Data. In: M. Treseler, editor, Statistics in a Nutshell: A Desktop Quick Reference. O'Reilly Media. p. 188-206.

Bruggmann, R., A.K. Bharti, H. Gundlach, J. Lai, S.Young, A.C. Pontaroli, F. Wei, G. Haberer, G. Fuks, C. Du, C. Raymond, M.C. Estep, R. Liu, J.L. Bennetzen, A.P. Chan, P.D. Rabinowicz, J. Quackenbush, W.B. Barbazuk, R.A. Wing, B. Birren, C. Nusbaum, S. Rounsley, K.F. Mayer, and J. Messing. 2006. Uneven chromosome contraction and expansion in the maize genome. Genome Res. 16(10):1241-1251.

Bureau, T.E., and S.R. Wessler. 1994. Mobile inverted elements of the Tourist family are associated with the genes of many cereal grasses. Proc. Natl. Acad. Sci. USA. 91:1411-1415.

Clark, R.M., S. Tavare, and J. Doebley. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Mol. Biol. Evol., 22(11):2304-2312.

Devos, K.M., J. Ma, A.C. Pontaroli, L.H. Pratt, and J.L. Bennetzen. 2005. Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc. Natl. Acad. Sci. USA. 102(52):19243-19248.

Erayman, M., D. Sandhu, D. Sidhu, M. Dilbirligi, P.S. Baenziger, and K.S. Gill. 2004. Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res., 32: 3546-3565.

Feschotte, C., N. Jiang, and S.R. Wessler. 2002. Plant transposable elements: where genetics meet genomics. Nat. Rev. Genet. 3**:**329-341.

Gu, Y.Q., D. Coleman-Derr, X.Y. Kong, and O.D. Anderson. 2004. Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. Plant Physiol. 135:459-470.

Guidet, F., P. Rogowsky, C. Taylor, W.Song, and P. Langridge. 1991. Cloning and characterization of a new rye-specific repeated sequence. Genome. 34:81-87.

Hua-Van, A., A.L. Rouzic, T.S. Boutin, J. Filee, and P. Capy. 2011. The struggle for life of the genome's selfish architects. Biol. Direct. 6(1):19-44.

Kalendar, R., A.J. Flavell, T.H.N. Ellis, T. Sjakst, C. Moisy, and A.H. Schulman. 2011. Analysis of plant diversity with retrotransposon-based molecular markers. Heredity. 106:520-530.

Kumar, A., K. Simon, M.J. Iqbal, M. Michalak de Jimenez, F.M. Bassi, F. Ghavami, O. Al-Azzam, T. Drader, Y. Wang, M.C. Luo, Y.Q. Gu, A. Denton, G.R. Lazo, S.S Xu, J. Dvorak, P.M.A. Kianian, and S.F. Kianian. 2012. Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genomics. 13:597-611.

Künzel, G., L. Korzun, and A. Meister. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics. 154(1):397-412.

Lehmensiek, A., W. Bovill, P. Wenzl, P. Langridge, and R. Apples. 2009. Genetic mapping in the Triticeae. In: C. Feuillet and G.J. Muehlbauer, editirs, Genetics and genomics of the Triticeae. Springer, New York. P. 201-235.

Li, W., P. Zhang, J.P. Fellers, B. Friebe, and B.S. Gill. 2004. Sequence composition, organization, and evolution of the core Triticeae genome. Plant J. 40(4):500-511. Metzker, M.L. 2010. Sequencing technologies-the next generation. Nature Reviews. 11:31-46.

Nasuda S, Y. Kikkawa, T. Ashida, A.K.M.R. Islam, K. Sato, and T.R. Endo. 2005. Chromosomal assignment and deletion mapping of barley EST markers. Genes. Genet. Syst. 80:357-366.

Paux, E., D. Roger, E. Badaeva, G. Gay, M. Bernard, P. Sourdille, and C. Feuillet. 2006. Characterization the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J. 48:463-474.

Rabinowicz, P.D., R. Citek, M.A. Budiman, A. Nunberg, J.A. Bedell, N. Lakey, A.L. O'Shaughnessy, L.U. Nascimento, W.R. McCombie, and R.A. Martienssen. 2005. Differential methylation of genes and repeat in land plants. Genome Res. 15(10):1431-1440.

SanMiguel, P.J., W. Ramakrishna, J.L. Bennetzen, C.S. Busso, and J. Dubkovsky J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct. Integrative Genomics. 2:70-80.

Sato, K., Y. Motoi, N. Yamaji, and H. Yoshida. 2011. 454 sequencing of pooled BAC clones on chromosome 3H of barley. BMC Genomics. 12:246-252.

Šimková, H., J.T. Svensson, P. Condamine, E. Hřibová, P. Suchánková, P.R. Bhat, J. Bartoš, J. Šafář, T.J. Close, and J. Doležel. 2008. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics. 9:294-302.

Sreenivasulu, N. A. Ganer, and U. Wobus. 2008. Barley genomics: An overview. Inter. J. plant genomics. 2008:1-13.

Suchánková, P., M. Kubaláková, P. Kovářová, J. Bartoš, J. Číhalíková, M. Molnár-Láng, T.R. Endo, J. Doležel. 2006. Dissection of the nuclear genome of barley by chromosome flow sorting. Theor. Appl. Genet. 113:651-659.

The international barley genome sequencing consortium. 2012. A physical, genetic and functional sequence assembly of the barley genome. Nature. 491:711-717.

Wanjugi, H., D. Coleman-Derr, N. Huo, S.F. Kianian, M.C. Luo, J. Wu, O. Anderson, and Y.Q. Gu. 2009. Rapid development of PCR-based genome specific repetitive DNA junction markers in wheat. Genome. 52(6):576-587.

Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morqante, O. Panaud, E. Paux, P. SanMiguel, and A.H. Schulman. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8(12):973-982.

Wicker, T., N. Stein, L. Albar, C. Feuillet, F. Schlagenhauf, and B. Keller. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J. 26(3):307-316.

Wicker, T., S. Taudien, A. Houben, B. Keller, A. Graner, M. Platzer, and N. Stein. 2009. A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J**.** 59:712-722.

You, F.M, H. Wanjugi, N. Huo, G.R. Lazo, M.C. Luo, O.D. Anderson, J. Dvorak, and Y.Q. Gu. 2010. RJPrimers: unique transposable element insertion junction discovery and PCR primer design for marker development. Nucleic. Acids. Res. 38:313-320.

4. CONCLUSION

In this study an RH map for barley chromosome 3H was developed. Compared to an analogous genetic map, the RH map had an average of 9.53-fold higher resolution, reaching to >262.40-fold better resolution in pericentromeric regions. The final RH map consisted of 113 markers spanning over 2577.4 cR and covered the overall chromosome length including poor-recombination regions. Map resolution of the RH map was approximately 2.22 Kb. This level of resolution enabled anchoring contigs as small as 1.50 Kb in size onto the RH map. We designed ~ 400,000 barley-specific RJMs distributed along chromosomes and determined that RJMs are ideal marker platform for genotyping barley RH populations. *In silico* analysis of RJ sequences enabled detection of 1,190,885 TEs in the barley genome and characterization of their distribution and insertion sites. The 3H-RH map and genome-specific markers developed in this study can instigate development of RH maps for the remaining barley chromosomes. Due to their high resolution and uniformity, RH maps are an ideal platform for aligning BAC contigs in assembly of the barley genome.

APPENDIX

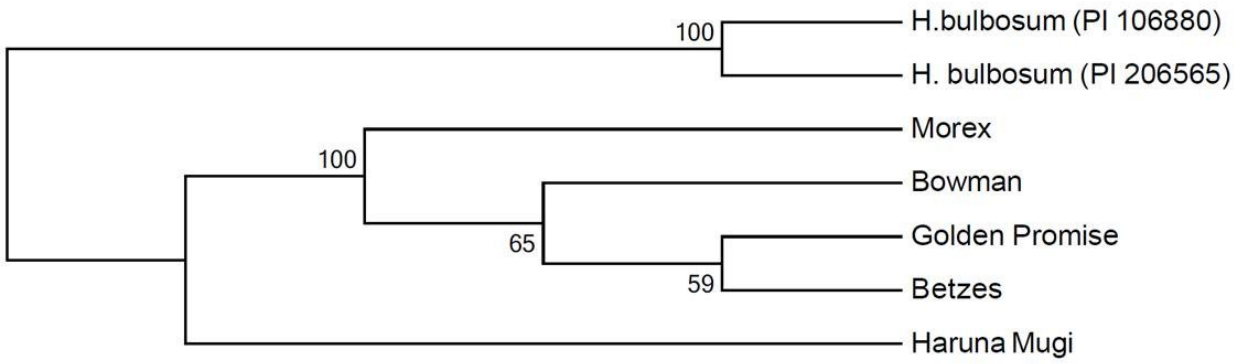

Figure A.1. Clustering seven barley genotypes with 33 RJMs. Five barley (*H. vulgare*) cultivars ("Betzes", "Golden Promise", "Bowman", "Morex", "Haruna Mugi"), and two wild barley (*H. bulbosum*) accessions PI 106880 (from Turkmenistan) and PI 206565 (from Greece) were genotyped by 33 RJMs. About 63% of the markers developed polymorphic amplicons, which separated all the seven lines. Clustering was performed by DARwin 5.0.158 software using NJ unweighted method.

	CACTA	*Copia*	*Gypsy*	*Harbinger*	*hAT*	*Helitron*	*Mariner*	*Mutator*	Unclassified Class I.	Unclassified LINE	Unclassified LTR	Unclassified SINE	Unclassified TIR
CACTA	3518 (1403.64)	590 (4661.00)	21494 (28795.14)	71(69.77)	3 (3.03)	2 (6.20)	2586 (615.63)	340 (100.64)	1 (1.36)	110 (38.44)	466 (187.51)	0 (0.45)	1792 (401.19)
Copia	1049 (2932.23)	12104 (9736.91)	62349 (60153.63)	27 (145.74)	4 (6.32)	11 (12.96)	39 (1286.07)	32 (210.24)	5 (2.85)	40 (80.30)	137 (391.70)	0 (0.95)	1 (838.10)
Gypsy	4225 (4547.36)	11361 (15100.19)	100980 (93287.40)	30 (226.02)	11 (9.81)	20 (20.10)	259 (1994.47)	109 (326.04)	3 (4.41)	65 (124.53)	482 (607.46)	0 (1.47)	4 (1299.74)
Harbinger	15 (21.12)	145 (70.14)	213 (433.31)	56 (1.05)	0 (0.05)	0 (0.09)	81 (9.26)	31 (1.51)	0 (0.02)	0 (0.58)	5 (2.82)	0 (0.01)	0 6.04)
hAT	2 (1.47)	5 (4.88)	29 (30.16)	0 (0.07)	2 (0.00)	0 (0.01)	0 (0.64)	0 (0.11)	0 (0.00)	0 (0.04)	0 (0.20)	0 (0.00)	0 (0.42)
Helitron	0 (0.97)	2 (3.21)	5 (19.84)	0 (0.05)	0 (0.00)	0 (0.00)	4 (0.42)	0 (0.07)	0 (0.00)	0 (0.03)	13 (0.13)	0 (0.00)	1 (0.28)
Mariner	32 (131.26)	266 (435.86)	1647 (2692.70)	30 (6.52)	0 (0.28)	8 (0.58)	1022 (57.57)	43 (9.41)	0 (0.13)	14 (3.59)	9 (17.53)	0 (0.04)	31 (37.52)
Mutator	58 (39.92)	121 (132.57)	481 (819.00)	244 (1.98)	0 (0.09)	0 (0.18)	19 (17.51)	102 (2.86)	0 (0.04)	5 (1.09)	1 (5.33)	0 (0.01)	1 (11.41)
Unclassified Class I	0 (0.81)	11 (2.70)	10 (16.67)	0 (0.04)	0 (0.00)	0 (0.00)	0 (0.36)	0 (0.06)	0 (0.00)	0 (0.02)	0 (0.11)	0 (0.00)	0 (0.23)
Unclassified LINE	17 (27.78)	204 (92.23)	461 (569.81)	1 (1.38)	0 (0.06)	0 (0.12)	10 (12.18)	1 (1.99)	0 (0.03)	19 (0.76)	1 (3.71)	3 (0.01)	1 (7.94)
Unclassified LTR	67 (135.09)	679 (448.58)	2603 (2771.27)	2 (6.71)	0 (0.29)	0 (0.60)	9 (59.25)	7 (9.69)	0 (0.13)	0 (3.70)	125 (18.05)	0 (0.04)	0 (38.61)
Unclassified SINE	1 (0.01)	0 (0.04)	0 (681.71)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.01)	0 0.00)	0 (0.00)	1 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Unclassified TIR	0 (33.23)	0 (110.35)	0 (681.71)	0 (1.65)	0 (0.07)	0 (0.15)	39 (14.57)	0 (2.38)	0 (0.03)	0 (0.91)	0 (4.44)	0 (0.01)	820 (9.50)

Table A.1. None-random associations among TE superfamilies within the barley genome. A total 10.22 Gb of the barlye genome 454 survey sequencung data were parsed with RJPrimers to detect TE-TE junctions. Numbers on the top are the observed values and numbers in paranthesis are the expected values of TE-TE associations.

Table A.2. The composition of TEs among seven barley chromosomes. Copy number of TE superfamilies was obtained based on the frequency of their corresponding junctions within the genome, using the script developed in this research. Composition of each TE superfamily was calculated based on the percentage of that superfamily copy number to the total number of detected TE superfamillies.

	1H	2H	3H	4H	5H	6H	7H
Retrotransposon	90.93%	86.97%	86.03%	86.85%	86.06%	87.32%	86.96%
LTR	90.66%	86.64%	85.71%	86.51%	85.73%	87.01%	86.65%
Gypsy	67.02%	64.73%	64.71%	64.69%	64.37%	65.51%	64.95%
Copia	23.65%	21.91%	21.00%	21.82%	21.37%	21.50%	21.70%
Unclassified LTR	0.68%	0.99%	1.00%	1.06%	1.01%	0.88%	0.90%
Non-LTR	0.27%	0.34%	0.32%	0.34%	0.33%	0.31%	0.31%
Unclassified LINE	0.27%	0.33%	0.32%	0.34%	0.32%	0.30%	0.31%
Unclassified SINE	0.00%	0.01%	0.01%	0.00%	0.01%	0.00%	0.00%
DNA transposon	8.39%	12.03%	12.97%	12.09%	12.92%	11.80%	12.14%
TIR	8.36%	11.99%	12.93%	12.05%	12.89%	11.76%	12.10%
CACTA	6.53%	9.43%	9.74%	9.18%	9.87%	9.37%	9.82%
Mariner	1.40%	1.52%	1.91%	1.70%	1.77%	1.43%	1.33%
Mutator	0.07%	0.37%	0.46%	0.44%	0.49%	0.41%	0.39%
hAT	0.01%	0.01%	0.01%	0.01%	0.02%	0.01%	0.01%
Harbinger	0.32%	0.41%	0.45%	0.39%	0.44%	0.32%	0.38%
Unclassified	0.02%	0.26%	0.35%	0.34%	0.31%	0.23%	0.17%
Non-TIR	0.03%	0.04%	0.04%	0.03%	0.03%	0.04%	0.04%
Helitron	0.03%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
Unclassified	0.00%	0.02%	0.02%	0.01%	0.01%	0.01%	0.01%
Total number of TEs	57981	186071	190290	172444	209542	190579	183978

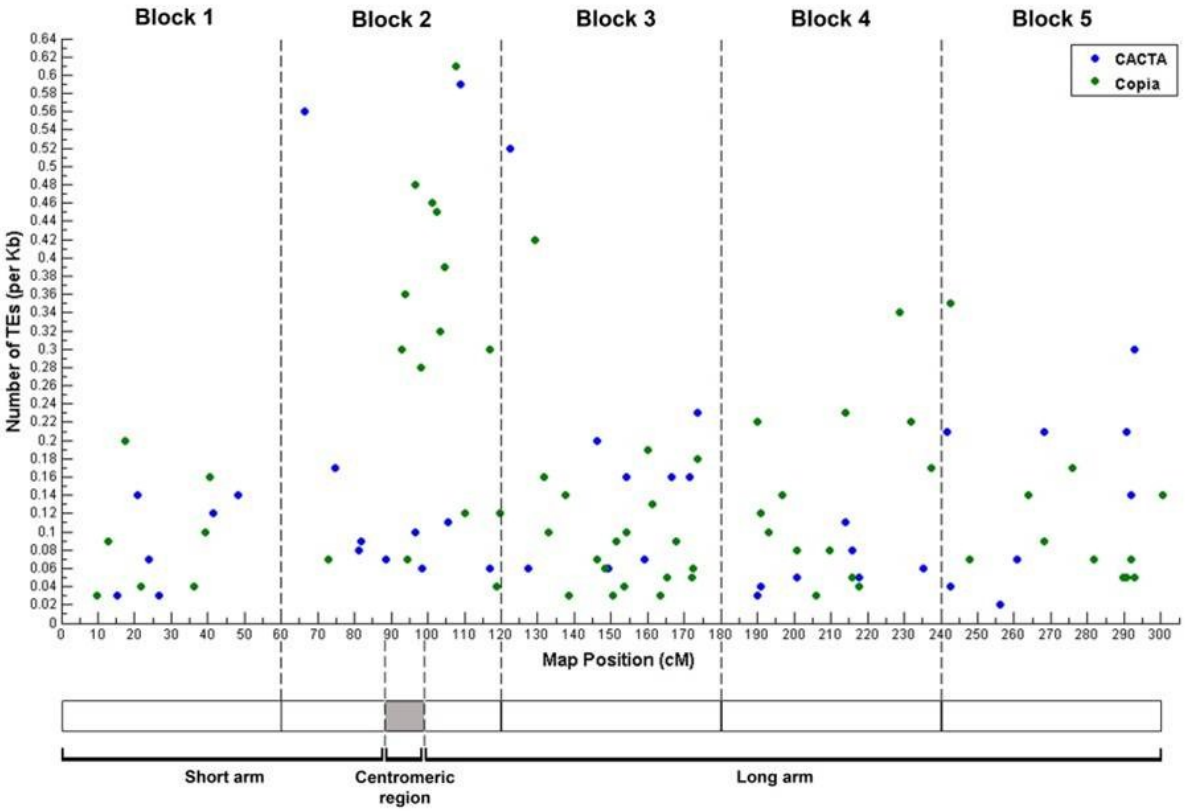

Figure A.2. Distribution of *CACTA* and *Copia* elements along the chromosome 3H. Chromosome 3H was divided into five blocks of 60cM intervals based on a genetic map developed by Sato et al. (2011). The number of *CACTA* and *Copia* elements per Kb sequences of contig anchored onto the genetic map was calculated. Centromeric region was identified based on markers flanking the centromere on a deletion bin map (Sakat et al., 2009

Table A.3. Testing the random distribution of transposable element superfamilies along the chromosome 3H. Chromosme 3H was divided into 60 cM-length blocls, based on a genetic map developed by Sato et al. (2011). The expected value of each TE superfamily was calculated by the observed number of that TE superfamily per 500 kb of all contigs anchored onto the genetic map and dividing the obtained value by five. The observed value of TE superfamilies within each block was calculated by the number of TEs per 500 kb of contigs anchored onto that block. Chi square test among the five blocks was performed at $p>0.001$ level of confidence.

	Expected	Block 1	Block 2	Block 3	Block 4	Block 5
CACTA	19.35	9.62	17.15	24.74	9.54	35.68*
Copia	31.06	11.37*	57.36*	25.47	35.77	25.35
Gypsy	36.66	19.25	48.25	47.31	46.90	21.59
Harbinger	6.71	8.75	7.50	6.55	7.95	2.81
hAT	2.62	0.53	0.00	0.00	0.00	0.00
Helitron	1.75	1.60	1.45	0.00	0.00	0.00
Mariner	35.63	30.625	34.85	40.03	34.18	38.50
Mutator	5.06	5.25	2.68	10.19	1.59	5.63
Unclassified class I	0.63	0.87	0.00	0.72	1.59	0.00
Unclassified LINE	10.07	2.62	15.54	14.55	6.36	11.26
Unclassified LTR	3.86	0.87	4.28	1.45	2.38	10.33
Unclassifie*d* SINE	0.21	0.00	1.07	0.00	0.00	0.00

Table A.4. The composition of RJs in gene-rich regions of the barley chromosome 3H. Number of RJs were obtained based on 172 gene-bearing contigs anchored onto the chromosome 3H (Sato et al., 2011).

	CACTA	*Copia*	*Gypsy*	*Harbinger*	*hAT*	*Helitron*	*Mariner*	*Mutator*	Unclassified DNA transposon	Unclassified LINE	Unclassified LTR	Unclassified SINE	Unclassified TIR	Non-TE
CACTA	3	1	0	0	0	0	0	0	0	0	0	0	0	205
Copia		10	8	1	0	0	1	0	0	0	0	0	0	345
Gypsy			7	0	0	0	1	1	1	0	0	0	0	389
Harbinger				0	0	0	0	0	0	0	0	0	0	83
hAT					0	0	0	2	0	0	0	0	0	3
Helitron						0	0	0	0	0	0	0	0	14
Mariner							4	0	0	0	0	0	0	413
Mutator								3	0	0	0	0	0	47
Unclassified DNA transposon									0	0	0	0	0	7
Unclassified LINE										2	0	0	0	141
Unclassified LTR											0	0	0	21
Unclassified SINE												0	0	4
UnclassifiedTIR													0	0

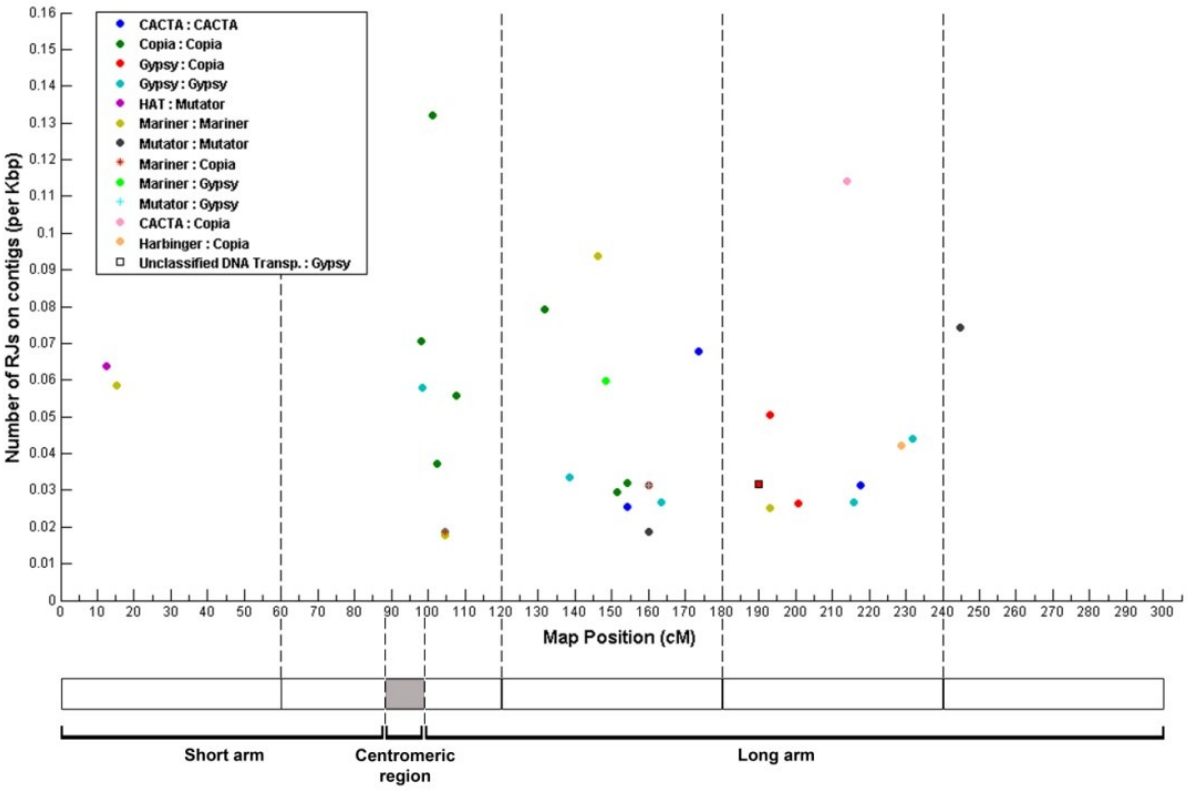

Figure A.3. Clustering of nested TE-TE junctions in pericentromeric and proximal regions of the long arm of barley chromosome 3H. Chromosme 3H was divided into 60 cM-intervals based on a genetic map developed by Sato et al.(2011). The number of "TE-TE" junctions were obtained based on Kb sequences of contigs anchored onto the chromosme 3H genetic map (Sato et al., 2011). Centromeric region was distinguished based on markers flanking the centromere on a deletion bin map (Sakat et al., 2009).